On the Volatility of Equity Prices*

Joseph M. Ostroy† Simon M. Potter‡

September 1994

(Preliminary version)

Abstract

We formulate a representative consumer model of intertemporal resource reallocation in which fluctuations in equity prices contribute to the smoothing of consumption flows. Features of the model include (a) an incompletely observable stochastic process of productivity shocks leading to fluctuating confidence of beliefs and (b) technologies involving commitments of a resource good. These features are exploited to show that (1) equities are not a representative form of total wealth and (2) the valuation of currently active firms is not representative of the valuation of all firms.

We examine the implications of (1) and (2) to argue that empirical findings for the volatility and 'value shortfall' of equity prices may be consistent with a frictionless representative consumer model having a low degree of risk-aversion. Simulation of a calibrated version of the model for a risk-neutral consumer shows that when the 'data' is analyzed according to current econometric procedures, it is found to exhibit volatility of the same order of magnitude as that found in the actual data, although the model contains no excess volatility

KEYWORDS: Equity premium, excess volatility, value shortfall, technological commitments, uncertainty of beliefs

JEL CLASSIFICATIONS: E13 (Neoclassical aggregative models), E44 (Financial market and the macroeconomy), G12 (Asset pricing), G14 (Information and market efficiency)

*Research supported in part by the National Science Foundation. Thanks to Alex David, Philip Johnson, Seongwhan Oh, and Kwanho Shin for helpful discussions. Geoffrey Gerdes provided invaluable research assistance.

†Economics Department, University of California, Los Angeles. Email: ostroy@econ.sscnet.ucla.edu

‡Economics Department, University of California, Los Angeles. Email: potter@econ.sscnet.ucla.edu

1 Introduction

Evidence from equity markets (LeRoy and Porter (1981), Shiller (1981), Mehra and Prescott (1985) among others) poses the following challenge: How can the prices of equities be so volatile and their rates of return be so high when their associated output flows are so smooth? Should this evidence be broadly interpreted as casting doubt on the 'fundamentals' view of asset pricing; should it be interpreted more narrowly as calling into question particular versions of those theories such as the frictionless representative consumer approach to asset pricing; or as we shall argue, is there a fundamentals, frictionless representative consumer model having the property that volatile equity prices, along with high rates of return, *contribute* to the smoothing of output and consumption?

We regard the challenge as the need to integrate two seemingly opposite views on asset pricing. One, a finance perspective, is that as efficient aggregators of dispersed information, asset prices should *be cushioned* against shocks in subsequent dividend flows. From the finance perspective, the behavior of equity prices appears especially dubious. The other, a price-theoretic perspective, is that equity prices act as 'shock absorbers:' the volatility of productivity shocks is transmitted directly to equity prices which, in turn, provide signals to the economy to reallocate resources to more productive uses *to cushion* the real impact of shocks on consumption. Evidently, from the so-called price-theoretic perspective the challenge appears less daunting. Integration of the two perspectives is a problem in general equilibrium theory, with the added restriction that there should be only one type of consumer.

The finance perspective is evident in the formulation of the challenge. For example, Shiller's *ex-post rational price* is an expression of the be-cushioned view of asset prices, as is Mehra and Prescott's calibration based on an exchange economy model. Further, the finance perspective has predominated in attempts to resolve the tension between the volatility of equities and the smoothness of dividends/consumption. To give a be-cushioned explanation of volatility, contributors have sought ways effectively to 'roughen' consumption flows. The smooth consumption stream can be turned into a more volatile stream of marginal utilities called stochastic discount factors by invoking risk-aversion (Grossman and Shiller (1981)), Hansen and Jagannathan (1991)) or different kinds of preferences (Epstein and Zin (1989)). But this route has proved problematic. The preferences required of the representative consumer for the behavior of equity prices to match consumption/dividends is much different from the preferences used to match the data in other applications of equilibrium business cycle models.

Instead of seeing the problem as understanding how to roughen consumption to make it fit the volatility of asset prices, we see the challenge as understanding how the volatility of equities helps to smooth consumption. The productivity shocks we introduce are such that *if* resources remained where they were, output flows would be volatile. However, by redirecting resources among firms the economy can adapt to the shocks in ways that minimize their consequences for aggregate fluctuations in dividends and consumption. The adaptation is guided by continually fluctuating equity prices.

The price-theoretic perspective suggests that in principle there is nothing contradictory between volatility of equity values and smoothness of consumption, particularly for the relation between individual or subsets of equities and aggregate consumption; however, for resource reallocation to contribute to the smoothing of aggregate consumption, the aggregate value of equities cannot fluctuate too much. By itself, the price-theoretic perspective does not appear capable of meeting the challenge posed by the empirical findings. Nevertheless, we shall argue that there is a further econometric implication of resource reallocation that makes equities appear to be more volatile than they are.

There are two different ways to show that reallocation is an issue which can make equities appear to be more volatile than they are. One is to assemble data on asset pricing which takes more explicit account of the problems posed by reallocation. An example of this approach is given below in our construction of 'year-by-year stock markets.' (See Section 2.2.) We demonstrate that after re-assembling the CRSP data on stock market prices to minimize problems caused by reallocation, one aspect of the equity premium called the 'value shortfall' is significantly reduced. Another way to address the problem, and the one emphasized here, is to analyze the consequences of (implicitly) accepting the hypothesis that reallocation is not an issue in a model where it is.

It is not unusual to assume that the econometrician does not have as much information as market participants. In our formulation, when the econometrician ignores the reallocation problem it is as if he is observing an economy with a (continually changing) set of incomplete equity markets, although they are actually complete. The econometrician (mis-)matches data from the real side of the economy on which output, consumption and dividends are produced via a complete markets model with data from incompletely observed equity markets. We shall examine the biases from such a procedure, both theoretically and through simulation. (Note: We rely on an ideal complete markets model to formalize the problem; however, a more general formulation is that whatever the extent of real-world incompleteness, the failure to take account of resource reallocation makes markets appear to be less complete than they are.)

Simulation is used to evaluate the magnitude of potential biases in favor of findings of excess volatility in an approach developed by West (1988). We find that an econometrician observing equilibrium asset prices and discount factors generated by our computational model calibrated to aggregate dividends in the U.S. would on average find excess volatility similar to that found by West given 100 years of data and even with 10,000 years of data would still think 50% of the volatility was excessive.

In the following section we describe in more detail the points of departure in our analysis of the volatility of equities. The consequences of two hypotheses, called representative wealth and representative firms, as well as implications of the failure of those hypotheses are highlighted. In Section 3 the model is described. We adopt the description of allocations as measures used in general equilibrium theory. In Section 4 some of the formal consequences of equities as non-representative wealth and the presence of non-representative firms are given.

These results are obtained via variational analysis of the optimal value function. In Section 5 we present a computationally tractable version of the theoretical model to measure the extent to which equities may appear to be more volatile than they are. There we point out biases in addition to the sampling problems associated with non-representative firms from employing linear forecasting models for dividends applied to the non-linear dividend generating process in our model. Finally, the data in Section 2.2 on the value shortfall for year-by-year stock markets is explained in the Appendix. (Appendix omitted in this version.)

2 The Representative Wealth and Representative Firms Hypotheses

The representative consumer approach to asset pricing follows Lucas (1978) and Brock (1982). Empirical applications frequently include one or more (implicit) restrictions that we call the representative wealth and the representative firm(s) hypotheses, each of which can be regarded as immediate extensions of Lucas' model to deal with production economies. The asset price anomalies cited above can be interpreted as rejections of the *joint* hypotheses of representative consumer/wealth or representative consumer/firm(s). Our goal is to show that there is an alternative representative consumer model satisfying neither added hypothesis which does not imply the strong restrictions on directly observable data frequently claimed for it, and that the model can produce smooth output and consumption while also generating equity valuations consistent with evidence from equity markets.

2.1 Equities Are Not Representative of Total Wealth: The Shock Absorber View

Well-known applications of representative consumer models to asset pricing include the implicit *representative wealth hypothesis* that equities are a representative source of total wealth. There is some indirect support for the hypothesis based on output flows: dividends from firms are approximately as smooth as 'dividends' from total wealth, namely consumption.

Calibrated versions of representative consumer *cum* representative wealth models find that equities command an excessively large return over short-term risk-free assets (Mehra and Prescott (1985)). These findings have been interpreted as challenging the validity of the representative consumer approach to asset pricing. The findings could also be interpreted as challenging the representative wealth hypothesis, particularly since corporate equities held directly by households are one of the most volatile components of total wealth. (See Figure 1).

In contrast to the representative wealth hypothesis, a standard feature of general equilibrium models is that there are two sources of wealth, one from resources and the other from the technologies employing them. Total wealth is therefore the sum of the valuations of resources plus the valuation of firms, where the latter represents future profits associated with those technologies (firms). In the model below, the wealth from firms is derived from their individual capacity constraints while the wealth from a single resource good is derived from its fixed supply, all according to standard principles of (stochastic) marginal productivity theory.

We shall see that Lucas' 'tree-model' is in our terms as if capacity constraints are the only ones that are binding, so that the *re*-allocation of the resource good to the various technologies is not a scarce factor. With capacity constraints as the sole form of wealth, it is necessarily representative. In the model we consider, capacity constraints and resources are both scarce; and it is fluctuations in their *relative scarcities* that is basic to our analysis of asset pricing.

As part of our goal to highlight the significance of the representative wealth and representative firm assumptions, we make two departures from the standard model: one is that within the Markovian class of productivity shocks we choose a particular family of processes exhibiting 'fluctuating confidence of beliefs' and the other is an extension from exchange to a production economy in which resources can be committed for varying lengths of time. These are complementary departures in the sense that the significance of one is dependent on the presence of other. (See Jones and Ostroy (1984).)

That there are two sources of wealth would mean little if their risk characteristics were the same. We formulate a representative consumer, but multi-sector technology, model in which wealth attributable to technologies — especially to technologies employing resources with long-term commitments — fluctuates more than wealth attributable to resources.

The aim of the model is to show that technological commitments, along with continually fluctuating confidence of beliefs, make firm valuations risky without transferring all of that risk to total wealth. The reason this is possible is that resource *re*-allocation among firms acts to (partially) absorb the stream of productivity shocks. Reallocation of resources has different implications for the payment streams to the two sources of wealth cooperating in its production. While reallocation of resources among firms acts to smooth aggregate consumption, the smoothing does not carry over to the aggregate valuation of the firms performing it. In the following paragraphs we give a heuristic explanation why.

The reallocation of resources away from firms likely to experience unfavorable productivity shocks toward those likely to have favorable shocks is based on Bayesian updating of beliefs. The belief process does not move precipitously from favoring one sector to another; rather it behaves more 'cautiously.' For example, if beliefs have supported the hypothesis that productivity shocks favor sector I (over II) in the past but recent evidence is not confirming, then beliefs will be 'less confident,' e.g., they are 50-50. Since beliefs represent an optimal forecast, it is therefore just as likely that future observations will favor I as that they will favor II. Hence, it would be a mistake to assume that because beliefs no longer favored I it was time to transfer resources to II. Rather, one should wait to see if future observations raise the beliefs about the likelihood that shocks are favoring II before making longer-term commitments. Consequently, resources are not immediately transferred from longer-term commitment technologies in one sector to longer-term commitments in another; instead, uncommitted resources are allocated to to shorter-term commitment technologies.

Total wealth decreases when beliefs exhibit more uncertainty, but only by a small amount since the increased uncertainty is transitory (although recurrent). The reduction in wealth is distributed asymmetrically. The valuation of each firm is, by construction, based on its firm-specific technology. Consequently, the value of *all* firms with long-term commitment technologies declines. Whether the anticipated resolution of uncertainty favors sector I or II affects the value of firms in those sectors more than its affects payments to the resource good, which will be used no matter how the uncertainty is resolved. Hence payments to the resource good hold their value as that component of wealth which captures the (well-founded) belief that the reduction in wealth due to increased uncertainty is only transitory. With increased uncertainty, payments to the resource good contain a return from their ability to store options.

When beliefs exhibit less uncertainty, the resource good also enjoys a high value because its marginal productivity is high; however, with decreased uncertainty its value does not rise by as much as the (longer-term) firm-specific technologies favored by the more certainly believed productivity shock. In other words, relative to the fluctuations in the aggregate value of firm-specific technologies, the value of the resource good falls less in periods of greater uncertainty (bad times) and rises less in periods of lower uncertainty (good times), causing a disproportionate fraction of the volatility in aggregate wealth to be born by the aggregate valuations of firms. Such a conclusion, however, goes only part of the way towards meeting the challenge of linking fluctuating equity values with smooth output and consumption. The remaining part is based on another mismatch between theory and data.

2.2 The Sampling Bias from Non-Representative Firms

Even if there is more than one source of wealth, the representative consumer necessarily holds it in representative form, i.e., as a representative portfolio of all firms and a representative amount of the resource good. The *representative firms hypothesis* is satisfied if there is a fixed set of firms from which output is always produced. Evidently, the representative consumer hypothesis does not imply representative firms. Output may be produced by a 'varying portfolio' of the currently active firms, i.e., the current winners. (For example, of firms originally in the CRSP index, only 10% remain.)

The same output and consumption — and therefore the same 'stochastic discount factors' — can be produced either by a model with fixed portfolio of representative firms or by a model with a varying portfolio of currently active firms. That depends on the (arbitrary) industrial structure assigning technologies to firms. For example, we could assume a single representative firm having a share of all technologies or, as we assume below, that each technology is a separate firm, and both would have the same output consequences.

Consider the difference in the market value of equities these two streams would imply. With a single representative firm whose total output is an aggregate of its many-plant technology there is no variation in individual firm valuations, whereas if each plant is a separate firm, there can be substantial variation among firm valuations. In both cases, however, there is a fundamental sense in which the same set of prices is guiding resource reallocation. It is just that with a single representative firm the variation of *internal* plant pricing remains unobserved, whereas when 'plants' are firms, plant pricing becomes market pricing.

Part of the market value of a firm comes from knowing when to de-activate some plants and activate others. No matter how subtle this problem may be, its implications will be reflected in the equity value of a representative firm. When plants are firms, firms will be continually disappearing and reappearing. For the representative consumer, the varying portfolio of currently active plants (or firms) within the entire portfolio is a more or less irrelevant detail, like a list of even-numbered equities. For the econometrician attempting to test the implications of an asset pricing model, the possibility that measured prices and returns derive from a varying portfolio can be important.

The difficulty in the non-representative firm environment is that observed equity values may no longer capture the consequences of inter-firm resource reallocation. For example, in the model below we assume that prices observed by the econometrician are only for active firms. Hence, the econometrician is not getting the same information from equity markets that would be present if firms were representative since temporarily de-activated firms have value. This assumption is intended to correspond to the phenomenon that firms enter stock markets, or indices based on them, when they are most active and leave when they are less active cease operations.

The representative firms assumption imposes tighter restrictions on commonly used financial data than a non-representative firms model. For example, to measure the value of a fixed group of representative firms over time it suffices at each date to examine the behavior of an index of firm activity such as the S&P 500, whereas if the composition of active firms is changing, the value of the index over time need not correspond to the values of any fixed group of firms.

If the tighter restrictions from the representative firms hypothesis are applied in a non-representative firm world, the econometrician will encounter a sampling bias from the mistaken conclusion that the risk characteristics of the current winners are representative of the entire population of firms. (Similar issues of measurement bias are found in Hopenhayn (1992) for Tobin's q and Olley and Pakes (1991) for productivities.) We shall argue that such qualifications have implications for the measurement of the equity premium and for the measurement of the volatility of equity prices.

Suppose dividend streams are produced by an intertemporally varying collection of firms. Then the (risk-adjusted) capitalized value of the stream will be *higher* than the observed firm valuations because the firms chosen to represent the production of future dividends, i.e., the current winners, are *not* necessarily the ones from which future dividends will be produced. Employing the representative firms hypothesis, the same evidence would be interpreted as the finding of a large premium on equities. For example, Hall and Hall (1993) document the existence of a 'value-shortfall' in the S&P 500: the value of firms in the index at a given date consistently underestimates the discounted value of dividends produced by firms in future indices. (See Figure 2.) Their interpretation is that the value shortfall reflects an excess discount rate that investors put on equities, while our explanation is that the value shortfall is evidence of continual reallocation to current winners.

Figure 3 illustrates that the value shortfall also holds for the CRSP data on stock market prices. We used the constant nominal discount factor of 5% in computing the 'ex-post rational price.'

From the (be-cushioned)/finance perspective the solution of the value shortfall is relatively straightforward: increase the degree of risk aversion of the representative consumer until the value shortfall disappears. Figure 4 shows the ex-post value of the marginal rate of substitution (stochastic discount factor) between tomorrow and today of a representative consumer, with constant relative risk aversion preferences, living in the U.S. from 1926 to 1992. Notice that as the degree of risk aversion increases the average value of the stochastic discount factor decreases well below one.

We then took a step beyond the Hall and Hall analysis to see if we could explain the value shortfall as the result of resource reallocation among a varying collection of firms. To isolate the consequence of resource reallocation, we extracted from the data the stream of future dividends produced by a given collection of firms at a given date. For example, using the firms listed in 1926, say a_{26}, b_{26}, c_{26}, etc. — called the 1926 stock market, we tracked their dividends from 1926 to 1992. In 1927, we performed the same procedure on the 1927 stock market of firms a_{27}, b_{27}, c_{27} etc., tracking their dividends from 1927 to 1992. Under the representative firms hypothesis, the computation of the value shortfall would be the same whether we regarded the stock market as a single entity over the sample period or as a succession of year-by-year stock markets. Under the varying portfolio of firms hypothesis, we would predict the value shortfall from the succession of year-by-year stock markets to be smaller than value shortfall from the stock market regarded as a single entity. Figure 5 reports the results of the test. It illustrates that a substantial portion of the value shortfall is eliminated when the ex-post rational price is computed from the succession of stock markets. In the Appendix, we describe how the data for the year-by-year stock markets and their associated dividends was assembled from the original data set. (...Appendix omitted..)

The failure of the representative firms hypothesis also has implications for the volatility of stock market indices. Unfavorable shocks to the current collection of firms produces a more significant decline in their valuations than will be evidenced by future dividends since the latter will be smoothed by reallocation from the current to future collections of firms so as to achieve efficient output and consumption smoothing. Similarly, favorable shocks to the current collection produces a more significant increase in their value than will be evidenced by future dividends since it implies that resource reallocation to another collection will be delayed.

Reallocation of resources among firms acts to smooth dividends in a way that is governed by the *population* of firms; but, in the absence of representative firms, econometric practice implies a (mis-)match between output, consumption or dividend data from the population of firms being imputed to valuations or returns from the (non-representative) *samples*. We shall argue that measured rates of return from the samples are biased in favor of finding an apparently excessive premium and measured prices of the sample are biased in favor of finding apparently excessive volatility, although returns and fluctuations in the value of the population obey standard properties.

3 The Model

3.1 Rationally Fluctuating Confidence in Beliefs

Productivity shocks will be determined below by random variables $\{Z_t\}$ that are a Markovian mixture of two processes $Z_t(1)$ and $Z_t(2)$:

$$Z_t = \lambda_t Z_t(1) + (1 - \lambda_t) Z_t(2).$$

Each $Z_t(i)$ is iid with $Z(1) \sim -Z(2)$. For example, if $Z(1)$ is a binomial random variable $[1, -1; p]$, then $Z(2) \sim [1, -1; (1-p)]$; or if $Z(1)$ is normally distributed, $Z(1) \sim \mathcal{N}(m, \sigma)$, then $Z(2) \sim \mathcal{N}(-m, \sigma)$.

Each λ_t can take on only two values, $\{0, 1\}$; hence, Z_t is determined either from $Z_t(1)$ or $Z_t(2)$. The two-state Markov transition for λ_t is symmetric. It is defined by the probability that the state at date t will repeat at $t+1$,

$$\mathbf{P}[\lambda_{t+1} = h \mid \lambda_t = h] = \alpha, \quad h \in \{0, 1\}.$$

We think of the *persistence parameter* α as near unity; hence, in the medium term Z_t will often be the outcome of successive draws from the same $Z(i)$; but over time the likelihood of a switch increases and in the long run the unconditional mean of Z is $\mathbf{E}Z = 1/2[\mathbf{E}Z(1) + \mathbf{E}Z(2)] = 1/2[\mathbf{E}Z(1) - \mathbf{E}Z(1)] = 0$. This particular conclusion holds for any $\alpha \in (0, 1)$, i.e., as long as switches from $Z(1)$ to $Z(2)$ are recurrent.

Assume that market participants know α and the properties of $Z(i)$ but observe only $\omega_t = (z_t, z_{t-1}, \ldots)$; hence, they can only make inferences from the sequence of realizations of Z about the sequence of distributions from which they were drawn. Let

$$b_t = \mathbf{P}[\lambda_t = 1 \mid \omega_{t-1}]$$

be the belief upon entering period t that the distribution from which Z_t will be drawn is $Z(1)$. $\mathbf{P}[\lambda_t = 1 | \omega_t]$ be the posterior on $\lambda_t = 1$ after observing Z_t. These posteriors are updated according to the following recursion

$$b_t = \mathbf{P}[\lambda_t = 1 \mid \omega_{t-1}] = \alpha \mathbf{P}[\lambda_{t-1} = 1 \mid \omega_{t-1}] + (1 - \alpha)(1 - \mathbf{P}[\lambda_{t-1} = 1 \mid \omega_{t-1}]).$$

Denote by b_Z the belief in the following period after observing Z and having belief b in the current period. For the (Borel) set B, $\mathbf{P}_i(b,B) = \int \mathbf{P}(b, b_Z \in B)\, d\mathbf{P}(Z = Z(i))$ is the probability that having beliefs b today will lead to beliefs in B tomorrow given that observations today are from $Z(i)$. The transition probability for the Markov scheme of beliefs can be shown to be

$$\mathbf{P}(b,B) = b\mathbf{P}_1(b,B) + (1-b)\mathbf{P}_2(b,B).$$

The equilibrium distribution of the Markov scheme is defined as

$$\mathcal{B}(B) = \int \mathbf{P}(b,B)\, d\mathcal{B}.$$

The symmetry of the two-state Markov chain on λ implies that $\mathbf{E}\mathcal{B} = 1/2$. The opposing symmetry of $Z(1)$ and $Z(2)$ implies that $\mathcal{B}$ is symmetric around its mean and that the distribution of $\mathcal{B}$ has two modes whose locations depend on the persistence parameter and the 'distance' between $Z(1)$ and $Z(2)$. (See Figure ?.)

Through Bayesian updating, the stochastic process $\{Z_t\}$ determines a sequence of beliefs $\{b_t\}$ exhibiting *fluctuating confidence or uncertainty*, i.e., episodes of beliefs near 1/2 followed by beliefs rising towards 1, then decreasing, etc. When b_t is near 1/2, confidence is at a minimum (uncertainty is at a maximum), while the closer b_t is to one (zero) the more confident or certain are beliefs that the true distribution at t is $Z(1)$ ($Z(2)$). Confident (more certain) beliefs are more likely to persist because the persistance parameter α is near one. For the same reason, less confident (more uncertain) beliefs will not be long-lasting, but they will be recurrent. Changes in beliefs represent learning, but information about future values of Z does not monotonically accumulate to some asymptote as it would if α were equal to unity. As beliefs move away from 0 or 1 toward 1/2, there is 'unlearning.'

Beliefs do not change precipitously. Suppose $b_t = b > 1/2$ and consider the possible paths $b_{t+\tau}$ 'hitting' a neighborhood of $b' < 1/2$ for the first time. If the distributions of $Z(i)$ are continuous, then a positive fraction of these paths will reach a neighborhood of 1/2 before they reach a neighborhood of b'. The closer b is to 1 and b' is to 0, the greater the probability that beliefs will pass through a neighborhood of 1/2 before reaching a neighborhood of b'. Thus, episodes of comparative certainty about $Z(1)$ are not typically followed by comparative certainty about $Z(2)$; rather, they are followed by a period of uncertainty. This uncertainty is justified since the sequence of observations which caused a change in beliefs from $b > 1/2$ to a neighborhood of 1/2 may have come from $Z(1)$. (See Figures 6–8.)

An application of the binomial version of this model to equity pricing is found in David et. al. (1992) and David (1994) gives a continuous-time version of the problem. Hamilton (1989) developed the econometric technology for estimating α in similar models.

3.2 A Model of Resource Reallocation with Commitments

The set of technologies is

$$F = \{f = (i,\ell,k) : i = I, II \quad \ell = 1 \ldots L \quad k \in [0,K]\ \} \cup \{f_o\}.$$

The index i refers to the sector, ℓ to the commitment ℓength of the technology and k to the '$k$apacity' of the technology to employ a resource good. The technology f_o will be called the option-storing technology. It belongs to neither sector, has the minimum commitment length of one period, and unlimited capacity to employ resources. Each technology is a type of firm.

Let $x = (x_0, x_1, \ldots, x_L)$ describe the allocation of a single resource good: x_0 represents the amount of the resource good committed today, x_1 the amount committed one period ago, x_2 the amount committed two periods ago, ..., and x_L the amount committed L periods ago. Also, define the sum of commitments as

$$\bar{x} = \sum_{n=1}^{L} x_n.$$

Note that $\bar{x}$ does not include the resources committed today. This reflects the following *timing convention*: decisions about resources committed today are made at the end of the period and do not come on line until tomorrow.

The pair $(f, x) \in F \times R_+^{L+1}$ describes an allocation to firm f. The set of feasible allocations for firm f is denoted by $X(f)$. If $f = (i, \ell, k)$ then

$$X(f) = \{x \in R_+^{L+1} : \bar{x} \leq k \ \& \ x_{\ell+1} = x_{\ell+2} \cdots = x_L = 0\}.$$

The *total* amount of resources committed in the previous ℓ periods cannot exceed k and commitments beyond those made ℓ periods ago have been released. If $f = f_o$,

$$X(f_o) = \{x \in R_+^{L+1} : x_2 = \cdots = x_L = 0\};$$

there are no capacity limits and commitments are for one period. The set of feasible firm allocation pairs is

$$C = \{(f, x) : x \in X(f)\}.$$

An *allocation* for the economy as a whole is described by a (positive) measure $a \in \mathbf{M}(C)$.

The total resources employed under a is

$$\int \imath_{\bar{x}} \, da,$$

where $\imath_{\bar{x}} \, da(f, x)$ represents the operation of summing the components of x (excluding x_0) to form $\bar{x}$ and then multiplying by the mass of firms of type f having allocation x. Loosely speaking, this is the integral of the second marginal of a.

The first marginal of a, denoted a_F, is a measure on the technologies in F, i.e., $a_F(G)/a_F(F)$ is the fraction of the firms in the economy having technologies in $G \subset F$.

Net investment associated with the pair (f, x) is $x_0 - x_{\ell(f)}$, the difference between the newly committed resources installed in f *at the end of the period* minus the resources released during the period. Aggregate net investment in the allocation a is

$$\int [\imath_{x_0} - \imath_{x_{\ell(f)}}] \, da.$$

When firm $f = (i, \ell, k)$ makes a (new) investment, it is always in technology (i, ℓ). Such a commitment of resources in one period will imply a corresponding commitment of resources in the following period. Define the set of resource commitments to firm f in the following period when its current commitments are x as

$$X(f, x) = \{x' : x' \in X(f) \;\&\; x_1' = x_0, x_2' = x_1, \ldots, x_\ell' = x_{\ell-1}\}.$$

Definition: An allocation a is *feasible* if

(i) (No growth in resources)

$$\int [\imath_{x_0} - \imath_{x_{\ell(f)}}]\, da \leq 0$$

The allocation a is assumed feasible throughout.

The allocation a' is a *feasible follower* of a if a' is feasible and

(ii) (No change is firms)

$$a_F' = a_F$$

(iii) (Individual feasibility of allocations) For $\mu = (a, a')$,

$$\mu\Big\{[(f, x), (f, x')] : x' \notin X(f, x)\Big\} = 0$$

Denote the set of feasible followers of a as $A(a)$.

(iv) The *sequence of allocations* $\{a_t\}_{t=1}^{\infty}$ is *feasible for* $a_0 = a$ if for $t = 1, \ldots$

$$a_t \in A(a_{t-1}).$$

Condition (i) says in the aggregate new commitments of resources do not exceed the resources about to be released; (ii) says that the distribution of firms in the follower allocation is the same as the distribution in the preceeding allocation; and (iii) says the set of individual allocations for firms in (a_t, a_{t+1}) that lie outside their feasible sets is null. Assuming (iii), equality in (i) implies that aggregate resources remain constant, i.e., $\int \imath_{\bar{x}}\, da = \int \imath_x\, da'$. We call the aggregate amount of resources w.

The key capacity restrictions of the model are (i) which limits the amount of resources available for employment and (iii) which limits the quantity of resources each firm can employ. If the capacity constraints k for each firm were unbounded, then (iii) would lose much of its effect as a binding constraint and technologies would not be scarce. In such a situation profits as the return to technologies would be zero and all the wealth would be imputed to the resource good. Conversely, if total resources were to exceed the sum of capacity constraints of all firms, then (ii) would not be binding and the resource good would effectively lose its scarcity. Consequently, all wealth would be attributable to technologies. For our purposes it is important to avoid both of these polar cases to emphasize fluctuations in the relative scarcities of technologies versus resources.

REMARK (LUCAS' EXCHANGE MODEL): The aggregate capacity constraints of firms, excluding the option-storing technology, is

$$\int_{F\setminus\{f_0\}} \imath_k \, da.$$

Assume the marginal product of resources applied to any firm is non-negative and ignore allocations to f_0 (assumed to be unproductive). Suppose the aggregate quantity of resources ew were at least as large as aggregate capacity. Then, reallocation would be pointless since each firm could achieve its capacity at no opportunity cost. Further, any interesting role for commitments or fluctuating confidence of the stochastic process of beliefs would be eliminated. Each firm could be regarded as an asset in an exchange economy producing a stochastic endowment from its fixed capacity, exactly as in Lucas (1978).

3.3 Productivity Shocks

Let $y_Z(f,x)$ be the output to firm f when its resource commitments are x and the productivity shock is Z. Assume output is linear in the input and the shock

$$y_Z(f,x) = s_Z(f)\bar{x},$$

where $s_Z(f)$ is the shock per unit of total resources committed to f in the event Z. For the special case of the option-storing technology f_o, $s_Z(f_o) = d \geq 0$. Therefore f_o is a riskless technology.

The shock $s_Z(\cdot)$ depends on f only through i and the commitment length of the firm, $\ell(f)$; so we write $s_Z(I,\ell)$. Recalling the definition of Z as determined either by $Z(1)$ or $Z(2)$ where $Z(1) \sim -Z(2)$, we transfer the symmetrically opposed values of $Z(i)$ to the sectors I and II by assuming

$$(A.1) \qquad s_Z(I,\ell) = s_{-Z}(II,\ell).$$

Letting $\mathrm{E}Z(1) > 0$, assume $s_Z(A,\ell)$ is increasing in Z. Therefore, draws from $Z(1)$ represent (on average) favorable shocks to firms in I, while draws from $Z(2)$ represent equally favorable shocks to firms in II. In terms of overall productivity there is no difference between firms in sectors I and II; the only distinction is how well-suited the firms are to respond to the current, but unknown, $Z(i)$.

The riskiness of a technology increases with its commitment length. For a $Z > 0$, assume

$$(A.2) \qquad s_Z(I,\ell+1) > s_Z(I,\ell) > s_{-Z}(I,\ell) > s_{-Z}(I,\ell+1).$$

Therefore the expected one-period ouput

$$\mathbf{E}_b(I,\ell) = b\int s_Z(I,\ell)\,d\mathrm{P}(Z=Z_1) + (1-b)\int s_Z(I,\ell)\,d\mathrm{P}(Z=Z_2),$$

($= \mathbf{E}_{(1-b)}(II,\ell)$) is increasing in b.

Recall that in the long run, $\mathbf{E}b = 1/2$. Assume that the long run expected one-period output from investing does not vary with commitment length,

$$(A.3) \qquad 1/2[s_Z(I,\ell) + s_{-Z}(I,\ell)] = 1/2[s_Z(I,\ell+1) + s_{-Z}(I,\ell+1)],$$

i.e., the higher output from guessing right ($Z > 0$ for I) is exactly offset by the lower output from guessing wrong. The higher expected one-period productivities associated with longer term commitment technologies are therefore based on their believed suitability to the current environment rather than on any intrinsic technological superiority.

If the parameter α determining the probability of a switch from $Z(1)$ to $Z(2)$ were set equal to $1/2$ so that there was no persistence, then $b_t = 1/2$ for all t. With beliefs constant at $1/2$, all technologies would have the same expected return. Hence, under risk neutrality, allocation of resources would be a matter of complete indifference. With risk aversion, however, resources would be applied only to one period commitment firms in I or II to minimize the risk associated with longer term commitments.

When α is near one, values of b_t near $1/2$ will occur, but beliefs are expected to change. In this situation there is an added reason for choosing short-term commitments — the option value associated with the fact that shorter term commitments are more flexible technologies. As beliefs move toward 1 (0), expected returns favor longer term commitments in I (II). By investing in shorter commitment technologies when b is near $1/2$, resources will be available for reinvestment in those longer term commitment technologies expected to be more productive in the future.

Aggregate output is always equal to aggregate consumption. Therefore aggregate consumption associated with the allocation a in the event of Z is

$$c_Z(a) = \int y_Z \, da = \int (s_Z \iota_{\tilde{z}}) \, da.$$

We shall be interested in how the reallocation process $\{a_t\}$ acts as a shock absorber on $\{s_{Z_t}\}$ to smooth aggregate consumption.

REMARK: Beliefs are relevant because they concern information about pay-off relevant productivity shocks, Z. Thus, the definition of fluctutating confidence or uncertainty in the previous section is not based on information such as changes in political events or general economic news having only an indirect payoff relevance. But to apply the model, we must assume the existence of economy-wide shocks interpreted by individual households and firms more or less uniformly as creating changes in confidence about the payoff relevance to them.

3.4 The Objective Function

The value function for the representative household model is

$$g_b(a) = \sup_{\{a_t\}} E_b \left\{ \sum_{t=0}^{\infty} \beta^t u(c_{Z_t}(a_t)) \right\},$$

subject to the conditions that $b_0 = b$ and $\{a_t\}$ is feasible sequence of allocations for $a_0 = a$. The one-period utility function $u : \mathbf{R}_+ \to \mathbf{R}$ is assumed to be concave and differentiable with derivative $Du(\cdot) > 0$ that is continuous at 0. This includes the possibility that u is linear.

It is readily established that the set of feasible allocations forms a compact and convex set and that the function g_b is concave.[1] Further, assuming as we shall that $\sup_Z s_Z(I, \ell)$ is bounded, $g_b(a)$ continuous in a, the given initial allocation. Therefore the sup exists.

Denote by a_Z the allocation chosen after observing Z when the allocation is a. Conditional on observations from $Z(i)$, the Markov transition is $\mathbf{P}_i[(a,b),(A,B)] = \int \mathbf{P}[(a,b),((a_Z,b_Z) \in A \times B)]\, d\mathbf{P}(Z = Z(i))$. The (unconditional) Markov transition can be shown to be

$$\mathbf{P}[(a,b),(A,B)] = b\mathbf{P}_1[(a,b),(A,B)] + (1-b)\mathbf{P}_2[(a,b),(A,B)].$$

An equilibrium distribution is

$$\mathcal{E}(A,B) = \int \mathbf{P}[(a,b),(A,B)]\, d\mathcal{E}$$

We do not examine conditions for which $\mathcal{E}$ is unique.

3.4.1 Directional Derivatives and the Option-storing Technology

With the assumptions above, the concave function $g_b(\cdot)$ can be shown to be Lipschitz. Consequently, the subdifferential $\partial g_b(a)$ can be shown to exist. An element of $\partial g_b(a)$ is a continuous function $q : C \to \mathbf{R}$ such that for all positive $\tilde{a} \in \mathbf{M}(C)$,

$$\int q\, d[\tilde{a} - a] \geq g_b(\tilde{a}) - g_b(a).$$

An element of the subdifferential 'prices' the initial allocation a. The inequality says that if a change from a to $\tilde{a}$ could be bought (or sold) at prices $q(f,x)$ for each (infinitesimal) firm of type f with resource commitments x, and if the consequences of the change were evaluated in terms of the overall consequences for utility, $g_b(\tilde{a}) - g_b(a)$, the cost of the change would exceed the utility gains.

In particular, if $a' = a + \delta_{(f,x)}$, where $\delta_{(f,x)}$ is the measure with unit mass at the point (f,x), the subdifferential inequality implies that

$$q(f,x) \geq Dg_b(a; \delta_{(f,x)}) \equiv \lim_{h \to 0_+} \frac{g_b(a + h\, \delta_{(f,x)}) - g_b(a)}{h}.$$

The directional derivative is well-known to be positively homogeneous, i.e., for $\gamma > 0$,

$$Dg_b(a; \gamma\delta_{(f,x)}) = \gamma Dg_b(a; \delta_{(f,x)}).$$

The expression $Dg_b(a; \gamma\delta_{(f,x)})$, in which an infinitesimal mass γ of firms each of type (f,x) is added to the economy should not be confused with $Dg_b(a; \delta_{(f,\gamma x)})$ in which a unit infinitesimal

[1] Compactness of the sequence of allocations is with respect to the topology for which $\{a_t^n\} \to \{a_t\}$ means that for each t, $a_t^n \to a_t$ in the sense of convergence in distribution.

mass of firms of type f with resources commitments γx is added. Indeed, γx may not even belong to X_f.

There is one situation, however, where the two directional derivatives do have the same value. Recall that the option-storing f_0 has no capacity constraint; hence $\gamma x \in X(f_o)$ whenever $x \in X(f_o)$.

The option-storing technology is a way to add or subtract resources from the model. If $e^0 = (1, 0, 0, \ldots, 0)$, $Dg_b(a : \delta_{(f_o, e^0)})$ represents the addition of an infinitesimal unit of resources that will become available for use, possibly with other technologies, from the following period onward. When $s_Z(f_o) = 0$, the value of the directional derivative represents the marginal productivity of resources (to be employed in other technologies) from the following period onward. It readily follows that

$$Dg_b(a; \delta_{(f_o, \gamma e^0)}) = Dg_b(a; \gamma \delta_{(f_o, e^0)}).$$

The equality also holds if $\gamma < 0$. In that case, the directional derivative represents the loss in future output from withdrawing an infinitesimal unit of resources from the following period onward. (See below.)

4 Variational Analysis of the Objective Function

4.1 The Division of Wealth between Technologies and Resources

In this section we apply marginal productivity theory to impute the total value of wealth to the firms and resources cooperating in its production. An obvious starting point is $Dg_b(a; \delta_{(f,x)})$. But this is composite marginal product consisting of the *joint* application of technology f along with resources x. The task will be to decompose $Dg_b(a; \delta_{(f,x)})$ into the sum of two components, one representing the value of the firm and the other representing the value of resources. The decomposition corresponds to a model of competitive markets for firms and resources. The flow version of this imputation requires that current output from any firm, $y_Z(f, x)$, must be imputed to the rental payments on resources plus the profits attributable to the technology. The latter term will consist of the competitively determined 'rents' on the technology plus a stochastic factor based on the composition of its commitments, i.e., luck.

DEFINITION: At the state (a, b), *total wealth* is

$$W(a, b) = \int q \, da,$$

where $q \in \partial g_b(a)$.

For any $q \in \partial g_b(a)$ it is well-known that $\int q \, da = Dg_b(a; a)$. Further,

$$Dg_b(a; a) \quad = \quad \lim_{h \to 0_+} \frac{g_b(a + h\,a) - g_b(a)}{h}$$

$$= \lim_{h \to 0_+} \mathbf{E}_b \left\{ \frac{\sum_{t=0}^{\infty} \beta^t [u(c_{Z_t}((1+h)a_t)) - u(c_{Z_t}(a_t))]}{h} \right\}$$
$$= \mathbf{E}_b \left\{ \sum_{t=0}^{\infty} \beta^t Du(c_{Z_t}(a_t))\, c_{Z_t}(a_t) \right\}$$

Recalling that $\omega_t = (z_t, z_{t-1}, z_{t-2}, \ldots)$ are the previous realizations, define

$$p_{\omega_t} = Du(c_{Z_t}(a_t))$$

as the stochastic component of the discount factor: it depends on the realization of Z_t and on ω_{t-1} through a_t. The flow version of total wealth is: for $q \in \partial g_b(a)$,

$$W(a,b) = \int q\, da = \mathbf{E}_b \left\{ \sum_{t=0}^{\infty} \beta^t p_{\omega_t} c_{\omega_t} \right\}.$$

In recursive form,

$$W(a,b) = \mathbf{E}_b\{p_Z c_Z(a) + \beta W(a_Z, b_Z)\}.$$

Recall that consumption equals output, $c_Z(a) = \int y_Z\, da$. To decompose total wealth into the portion attributable to firms and the portion attributable to resources, we shall decompose the output flow for each firm $y_Z(f,x)$ into profits and payments to resources as

$$y_Z(f,x) = \pi_Z(f,x) + r\bar{x}.$$

Note that payments for new commitments x_0 are not included in $\bar{x}$; they are paid for in the following period.

Spot markets operate as follows: at the start of period t, before Z_t is observed, the market price r_t for the resource good is established; after Z_t is observed, the market price p_t for the consumption/output good is determined. Because the firm must honor its previous commitments, firm (f,x) has an inelastic demand $\bar{x}$ for the resource good. Unless the firm will be capacity constrained in the following period, i.e., when $f = (i, \ell, k)$, unless $\sum_{h=1}^{h=\ell-1} x_i = k$, it also has an elastic demand for new commitments x_0 based on what the firm believes market prices for the output and resources goods will be in the future.

To give a marginal productivity interpretation of rental prices for resources, assume that the payoff to the riskless option-storing technology f_o is 0. Thus, $-Dg_b(a; \delta_{(f_o, -e^0)})$ is the marginal product of losing resources; more specifically, it represents the utility losses (which are zero in period 0) from having to withdraw a small amount of resources each period. For this to be well-defined, $\int \imath_{x_{t(f)}}\, da_t > 0$ for $t = 0, 1, \ldots$; otherwise, if at some t no resources were released from commitment, it would not be feasible to withdraw resources.

The market price of the resource good will determine a *marginal firm* as one that is just indifferent between making and not making new commitments. If (new) commitments were for one period only, a marginal firm in period t would one with the smallest expected per unit output, $\min_f\{\mathbf{E}_b s_{Z_t}(f) : (f,x) \in \operatorname{supp} a_t \ \&\ x_0 > 0\}$. With longer commitments, determination of the current marginal firm requires looking ahead to the implications of new commitments.

Let $e^n \in \mathbf{R}^{L+1}$ be the vector with a one in the n^{th} component ($n = 0, 1, \ldots, L$) and zeroes elsewhere and denote by e an abitrary element of $\{e^n\}$. Define $\Gamma(f, e)$ as the feasible successor(s) of (f, e) as follows: if $n = 0, \ldots, \ell(f) - 1$, $\Gamma(f, e^n) = (f, e^{n+1})$—resource commitments remain in the same firm and are aged; and (2) if $n = \ell(f)$, $\Gamma(f, e^n) = \{(f', e^0) : f' \in F\}$—released resources can be transferred to a new commitment in any firm.

Proposition 1 *Assuming resources are released in each period,*

$$-Dg_b(a; \delta_{(f_o, -e^o)}) = \min_{f_t} \mathbf{E}_b \left\{ \sum_{t=1}^{\infty} \beta^t p_{\omega_t} s_{Z_t}(f_t) \bar{e}_t \right\}$$

subject to the restriction that $(f_{t+1}, e_{t+1}) \in \Gamma(f_t, e_t), t = 1, 2 \ldots,$ where f_t is selected from those firms in $\{f : (f, x) \in \operatorname{supp} a_t, x_0 > 0\}$ in which new commitments are being made.

Letting $\{f^*_{\omega_t}\}_{t=1}^{\infty}$ be the sequence of firms from which resources are withdrawn that achieves the above minimum, then

$$-Dg_b(a; \delta_{(f_o, -e^o)}) = \mathbf{E}_b \Big\{ p_Z s_Z(f^*_Z) - \beta Dg_{b_Z}(a_Z; \delta_{(f_o, -e^o)}) \Big\},$$

where f^*_Z is the marginal firm in period 1. Therefore, setting

$$r_Z = s_Z(f^*_Z),$$

as the price of resources in period 1 after observing Z implies that the expected profits for the marginal firm will be zero.

An equally valid measure of the resource good is $Dg_b(a; \delta_{(f_o, e^o)})$; but there is a possible difficulty in equating the two. Any (f, x) that is capacity constrained is unwilling to pay for more of the resource good today because its marginal product would be zero. Therefore, if $f^*_{\omega_t}$ happens to be capacity constrained, there will be a difference between the cost of taking resources away and the gains from adding resources *for that firm.* Of course, for the economy as a whole the value of additional resources will be based on giving them to some other firm that it not capacity constrained; but we cannot be sure when the switch is made that the marginal product will not be lower, with the consequence that $-Dg_b(a; -\delta_{(f_o, e^o)}) > Dg_b(a; \delta_{(f_o, e^o)})$.

We shall ignore the possibility of a kink by assuming that the marginal firm is not capacity constrained. Therefore, the resource component of the value of total wealth at (a, b) is

$$\begin{aligned} R &= mw \\ &= \Big[p_Z r_0 + Dg_b(a; \delta_{(f_o, e^o)}) \Big] w \\ &= \left[p_Z r_0 + \mathbf{E}_b \left\{ \sum_{t=1}^{\infty} \beta^t p_{\omega_t} r_{\omega_{t-1}} \right\} \right] w, \end{aligned}$$

where $p_Z = p_{Z_0}$ and r_0 is the price of resources in the initial period.

Pricing the resource good yields an evaluation of firms. The stream of profits accruing to firm f starting at x is

$$\begin{aligned} v(f,x) &= \mathbf{E}_b\left\{\sum_{t=0}^{\infty}\beta^t p_{\omega_t}\pi_{Z_t}(f,x_t)\right\} \\ &= \mathbf{E}_b\left\{\sum_{t=0}^{\infty}\beta^t p_{\omega_t}[y_{Z_t}(f,x_t) - r_{Z_{t-1}}\bar{x}_t]\right\} \\ &\geq \mathbf{E}_b\left\{\sum_{t=0}^{\infty}\beta^t p_{\omega_t}[y_{Z_t}(f,x'_t) - r_{Z_{t-1}}\bar{x}'_t]\right\} \end{aligned}$$

for all $\{x'_t\}$ such that $x'_t \in X(f)$ and $x'_{t+1} \in X(f,x'_t)$ and $x'_0 = x$. Or, in recursive form,

$$v(f,x) = \max_{x' \in X(f,x)} \mathbf{E}_b\{p_Z\pi_Z(f,x) + \beta v(f,x')\}$$

Therefore the firms' component of the value of total wealth at (a,b) is

$$V = \int v\, da.$$

Current realized profits are

$$\pi_Z(f,x) = y_Z(f,x) - r\bar{x},$$

which can be positive or negative. For example, firms in sector 1 which took on long-term commitments expecting shocks to favor them may find that beliefs now favor firms in sector 2, therefore lowering $\mathbf{E}y_Z(f,x)$ from what it was when the commitments were taken on, without reducing the overall demand for resources, therefore keeping r high. Or, firms in sector 1 which took on commitments because the expected profit seemed to be just worth it may now find that prospects are even more favorable, raising $\mathbf{E}y_Z(f,x)$ by more than any expected increase in r.

To summarize,

Proposition 2 *Assuming resources are released in each period and the marginal firm is not capacity constrained,*

$$\begin{aligned} W &= \int q\, da \\ &= \mathbf{E}_b\left\{\sum_{t=0}^{\infty}\beta^t p_{\omega_t}\int y_{Z_t}\, da_t\right\} \\ &= \mathbf{E}_b\left\{\sum_{t=0}^{\infty}\beta^t p_{\omega_t}\left[\int \pi_{Z_t}\, da_t + r_{\omega_{t-1}}\int \imath_{\bar{x}}\, da_t\right]\right\} \\ &= \mathbf{E}_b\left\{\sum_{t=0}^{\infty}\beta^t p_{\omega_t}\int \pi_{Z_t}\, da_t\right\} + \mathbf{E}_b\left\{\sum_{t=0}^{\infty}\beta^t p_{\omega_t} r_{\omega_{t-1}}\int \imath_{\bar{x}}\, da_t]\right\} \\ &= \int v\, da + mw \\ &= V + R \end{aligned}$$

The term $q(f,x)$ measures the value of adding a firm with technology f *along with* the addition of (committed) resources x, whereas

$$v(f,x) = q(f,x) - r\bar{x},$$

measures the value of the *firm's* contribution to that combined value by subtracting off a measure of the economy-wide opportunity cost of x. But because it also depends on x, $v(f,x)$ remains a composite value that should be distinguished from $v(f,0)\,(= q(f,0))$. The latter is more nearly a measure of the 'pure' value of adding an additional firm with technology f. When g_b is differentiable, $v(f,0) = Dg_b(a;\delta_{(f,0)})$; therefore $v(f,0)$ is the value as of today from adding a firm with technology f which can be activitated tomorrow.

REMARK: (Bankruptcy) Evidently, since firms do not have to be used $v(f,0) \geq 0$; also the aggregate value of firms V will be non-negative. But we cannot rule out the following 'bankruptcy condition:' for states in the limiting distribution $\mathcal{E}$, we may encounter $v(f,x) < 0$.

4.1.1 Special Cases of the Marginal Firm

When the riskless payoff d from the option-storing technology f_o is zero, it may never be employed in an optimal sequence of allocations. Let d^* be the smallest value of d such that it is employed in every allocation in the support of an equilibrium distribution $\mathcal{E}$. In that case, it is readily seen that $-Dg_b(a;-\delta_{(f_o,e^0)}) = Dg_b(a;\delta_{(f_o,e^0)})$ and the value of the resource good is always $r_t = d$. Hence,

$$R = \mathbf{E}_b\left\{\sum_{t=0}^{\infty}\beta^t p_{\omega_t}\right\} w.$$

In addition to d^*, suppose the representative consumer is risk-neutral. Setting $p_{\omega_t} = 1$,

$$R = \frac{d^*}{1-\beta} w.$$

Further, since consumption equals output,

$$c_{Z_t}(a_t) = \int \pi_{Z_t}\, da_t + d^* w.$$

Therefore, all of the volatility in consumption is attributable to the volatility of aggregate profits. We shall consider these special cases below and throughout Section 4.

4.2 The Difference between Equities and Total Wealth

One way to compare equities with total wealth is to consider their coefficients of variation. The mean and standard deviation of total wealth are given by

$$\mathbf{E}W = \int W\, d\mathcal{E}; \qquad \sigma W = \left(\int [W - \mathbf{E}(W)]^2\, d\mathcal{E}\right)^{1/2}.$$

with similar expressions for the mean and standard deviation for V.

We are interested in conditions for which

$$\frac{\sigma V}{\mathbf{E}V} > \frac{\sigma W}{\mathbf{E}W}.$$

It is useful to rewrite the inequality as,

$$\frac{\sigma V}{[\sigma^2 V + \sigma^2 R + 2\mathbf{cov}(V,R)]^{1/2}} > \frac{\mathbf{E}V}{\mathbf{E}W}.$$

The following conditions contribute towards making equities more volatile than total wealth:

- $\frac{\mathbf{E}V}{\mathbf{E}W}$ is small
- σR is small compared to σV
- $\mathbf{cov}(V,R)$ is negative

Empirically, equities held by households represent no more than one-tenth total wealth and the evidence from Figure 1 suggests that it is more volatile than any other source of wealth. Information on the covariance among different sources of wealth has not yet been obtained. In its absence, we consider covariance from the perspective of the model.

Suppose the value of one-period output from the riskless technology is d^* of the previous section. Therefore

$$\begin{aligned} R &= \mathbf{E}_b\left\{\sum_{0=1}^{\infty}\beta^t p_{\omega_t}\right\}(d^* w) \\ &= \Gamma(d^* w) \end{aligned}$$

Therefore,

$$\mathbf{cov}(V,R) = (d^* w)\mathbf{cov}(V,\Gamma).$$

One should expect that $\mathbf{cov}(V,W) > 0$ whenever firms other than the option-storing technology are used. And, the greater the value of total wealth, the greater the consumption and therefore the lower the values of discounted p_{ω_t}'s; hence the lower the value of Γ, implying $\mathbf{cov}(\Gamma,W) < 0$. Therefore, $\mathbf{cov}(V,\Gamma)$ is negative.

When there is also risk-neutrality, Γ is a constant equal to $[1-\beta]^{-1}$ and the above inequality is always satisfied.

Another way to distinguish equities from total wealth is to compare their rates of return. It is convenient to begin with the expression for total wealth at the state (a_Z, b_Z),

$$W(a_Z,b_Z) = \mathbf{E}_{b_Z}\left\{p_{Z'}c_{Z'}(a_Z) + \beta W(a_{Z'},b_{Z'})\right\},$$

where Z' is the shock following Z. Use this to define

$$W_Z(a_Z,b_Z) = \frac{\beta}{p_Z}W(a_Z,b_Z),$$

as the 'ex-dividend' value of total wealth.

The predicted gross rate of return on wealth is

$$\varrho_W(a,b) = \mathbf{E}_b\{\varrho_{W_Z}(a_Z,b_Z)\},$$

where

$$\varrho_{W_Z}(a_Z,b_Z) = \mathbf{E}_{b_Z}\left\{\frac{W_{Z'}(a_{Z'},b_{Z'}) + c_{Z'}(a_Z)}{W_Z(a_Z,b_Z)}\right\}$$

The mean rate of return on total wealth is

$$\mathbf{E}\varrho_W = \int \varrho_W \, d\mathcal{E}$$

Similar definitions apply to $V_Z(a_Z,b_Z) = (\beta/p_Z)V(a_Z,b_Z)$, $R_Z(a_Z,b_Z) = \beta/p_Z R(a_Z,b_Z)$, etc., leading to a mean rate of return on equities $\mathbf{E}\varrho_V = \int \varrho_V \, d\mathcal{E}$ and on resources $\mathbf{E}\varrho_R = \int \varrho_R \, d\mathcal{E}$.

Equilibrium implies that for each b

$$\mathbf{E}_b\left\{\frac{\beta p_{Z'}}{p_Z}[\varrho_V - \varrho_W]\right\} = 0.$$

Because the equality holds unconditionally, after substituting it may be rewritten as

$$\mathbf{E}\left(\frac{\beta p_{Z'}}{p_Z}\right)\mathbf{E}[\varrho_V - \varrho_W] + \mathbf{cov}(\frac{\beta p_{Z'}}{p_Z}, \varrho_V - \varrho_W) = 0.$$

High values of the stochastic discount factor $\beta p_{Z'}/p_Z$ are associated with falling consumption, i.e., the existing deployment of committed resources is now less productive. Typically, this will occur because an unexpected switch has taken place between $Z(1)$ and $Z(2)$ causing beliefs to become less confident. Consequently the rate of return on V falls. But the rate of return on W is held up by the resource component of total wealth which is expected to be efficiently reallocated. The negative covariance between $\beta p_{Z'}/p_Z$ and $[\varrho_V - \varrho_W]$ implies that the rate of return on equities will exceed the rate of return on total wealth.

4.3 The Division of Wealth between Active and Inactive Firms

DEFINITION: Let $a = a_A + a_{\sim A}$, where a_A is the restriction of a to the set

$$\{(f,x) : \bar{x} \neq 0\},$$

the active firms with resource commitments, and $a_{\sim A}$ is the restriction of a to the complement

$$\{(f,x) : \bar{x} = 0\}),$$

the inactive firms. Since the allocation $a_{\sim A}$ is supported on $\{(f,0) : f \in F\}$, it is effectively a measure $(\tilde{a})^F \leq a^F$ on the available technologies not currently in use. We shall be particularly interested in those situations where the supports of a_A and $a_{\sim A}$ differ, reflecting the fact that different technologies are in use at different times.

Use this decomposition of a to define a decomposition of firm valuations

$$V = V_A + V_{\sim A} = \int v \, da_A + \int v \, da_{\sim A},$$

where V_A is the aggregate value of active firms and $V_{\sim A}$ is the aggregate value of inactive firms. It should be evident that $(f, 0) \in \operatorname{supp} a_{\sim A}$ does *not* imply that $v(f, 0) = 0$. On the contrary, $v(f, 0)$ is likely to be positive if there is a positive probability that f will be activated at some future date.

By construction all of the output and profits are produced by active firms, i.e.,

$$\int \pi_Z \, da = \int \pi_Z \, da_A \qquad \int y_Z \, da = \int y_Z \, da_A$$

Define the *value shortfall* as $V - V_A$, the difference between the value of all firms and the value of active firms.

It is important to distinguish who will observe the value shortfall and how it will be interpreted. First, households in our model do not observe a shortfall since they hold all the firms in V. The valuation of inactive firms makes them indistinguishable from the active ones as far as management of the household's portfolio of firms is concerned.

Taken literally, quotation of valuations for firms that will only be active in the future is a considerable idealization. More realistically, think of the active firms as those used in a stock market index. Firms may leave the index (become inactive) for a variety of reasons including a temporary change in the composition of the index, the firm leaves the stock market, or the firm ceases production. Conversely, firms become active by moving onto the index when they are already on the stock market, or when can change from privately to publicly held, or when new firms appear on the stock market, e.g., through mergers.

Our main interest is the econometrician's observation of the shortfall.

ASSUMPTION: The econometrician only observes the valuation of active firms.

There are situations where the econometrician will nevertheless be in a position similar to the household.

DEFINITION: Regard elements of F as 'plant' technologies. Given a_F as the distribution of plants in the economy, let a firm of type $\tilde{f}_j$ be defined as having the distribution of plants a_j^F. The *representative firms hypothesis* is satisfied if for all j, $\operatorname{supp} a_j^F = \operatorname{supp} a^F$. The firm sector of the economy is given by weights w_j, with $w_j > 0$ and $\sum w_j = 1$, such that $\sum w_j a_j^F = a^F$.

With respect to the firms $\tilde{f}_j$, the allocation a is summarized by a_j, where $(a_j^F) = a_j^F$ and $\sum w_j a_j^F = a$. The valuation of $\tilde{f}_j$ is based on the plants owned by $\tilde{f}_j$ as well as the resources committed to those technologies. The internal valuation of these plants in given by $v(f, x)$, above. Therefore the value of firm $\tilde{f}_j$ is $V_j = \int v\, da_j$. The value of all firms is therefore

$$\sum_j w_j V_j = \sum w_j \int v\, da_j = V.$$

Consequently with representative firms there is never a value shortfall: *future reallocations are fully reflected in the valuations of active firms.*

The definition adopted in this paper is at the opposite extreme from representative firms. Identifying firms with 'individual plants' makes them sufficiently heterogeneous that — along with the assumptions of commitments and fluctuating confidence of beliefs — future reallocations are *not* fully reflected in the valuations of active firms.

With non-representative firms, the observations of the econometrician and the household differ. Throughout the following assume that the econometrician fails to take this into account as if the representative firms hypothesis is assumed, although it is false.

When data on V_A is matched with the stream of future dividends, the value shortfall leads to an equity premium. From the model, we know that

$$V = \mathbf{E}_b \left\{ \sum_{t=0}^{\infty} \beta^t p_{\omega_t} \Pi_{\omega_t} \right\}.$$

where $\Pi_{\omega_t} = \int \pi_{Z_t}\, da_t$.

Observing V_A and dividends from a stream of realized profits $\{\Pi_{\omega_t}\}$, the econometrician looks for discount factors p'_{ω_t} such that

$$V_A = \sum_{t=0}^{\infty} \beta^t p'_{\omega_t} \Pi_{\omega_t}.$$

Hence, for the typical realization of $\{\omega_t\}$, the average value of p'_{ω_t} will be lower than $\mathbf{E}p_{\omega_t}$, leading to the conclusion that the observed discount rate on equities is larger than it actually is.

Replacement of V by V_A also has implications for measured volatility. Taking the coefficient of variation as a measure of volatility, the volatility of V_A is greater than V if

$$\frac{\sigma V_A}{\left[\sigma^2 V_A + \sigma^2 V_{\sim A} + 2\mathrm{cov}(V_A, V_{\sim A})\right]^{1/2}} > \frac{\mathbf{E}V_A}{\mathbf{E}V}$$

In the model, the right-hand side is less than one but is unlikely to be as low as 1/2. Empirically, however, the value of equities in a stock market index is likely to be substantially less than the value of all corporate equities. To mimic such a feature in this model, we could choose a selected subset of active firms — say those with commitment length greater than ℓ^* and define that as V_A. Such a modification would appear to increase the value of the right-hand-side over the left.

Returning to the original definition of V_A, the nearer the persistence parameter α is to one, the closer V_A/V will be to one. The left-hand side says the inequality will be greater

(a) the smaller the variance of $V_{\sim A}$ relative to V_A and (b) the smaller the covariance between them. $\sigma^2 V_{\sim A}$ may be smaller than $\sigma^2 V_A$ because the former is not contaminated by the presence of commitments. The value of inactive firms includes the option of being able to activate them which should act as a stabilizing influence on the total value of firms.

An expression of the shock absorber theme is that the covariance between active and inactive firms may be negative. The value of active firms increases beyond its mean when beliefs are fairly confident and resources are allocated consistently with those beliefs. In that case, the option value of inactive firms is worth less and therefore $V_{\sim A}$ will be below its mean. Similarly, V_A will be below its mean when beliefs are uncertain and resources are on average allocated to firms with shorter term commitments. In this situation there is an advantage to being inactive.

We also compare expected returns. As in the comparison between equities and total wealth, equilibrium implies that for each b,

$$\mathbf{E}_b \left\{ \frac{\beta p_{Z'}}{p_Z} [\varrho_{V_A} - \varrho_{V_{\sim A}}] \right\} = 0,$$

and therefore

$$\mathbf{E}\left(\frac{\beta p_{Z'}}{p_Z}\right) \mathbf{E}[\varrho_{V_A} - \varrho_{V_{\sim A}}] + \mathbf{cov}(\frac{\beta p_{Z'}}{p_Z}, \varrho_{V_A} - \varrho_{V_{\sim A}}) = 0.$$

A large value of the stochastic discount factor $\beta p_{Z'}/p_Z$ means consumption is falling. The unfavorable productivity shocks imply that dividends paid by active firms go down. Typically, this be associated with a change in beliefs so that resources are no longer properly deployed among the currently active firms. Some of these active firms will be continually contracting. Hence, the rate of return on active firms is falling. As another manifestation of the shock absorber role, the rate of return on inactive firms will rise because the fall in consumption is typically a signal that the currently inactive firms will have a greater role to play in the near future.

5 Calibration

The model described above has a number of features making its numerical solution and subsequent calibration impractical. For simulations and econometric work we propose a simpler version that captures the concepts of non-representative wealth and non-representative firms. The computational model leads to a calibration similar in form to Mehra and Prescott (1985): match unconditional moments of output, dividends and consumption and examine their pricing implications. Our goal is to illustrate the magnitude of the bias on the finding of apparent excess volatility when the implications of resource reallocation by non-representative firms is ignored. These implications extend to the appropriate choice of statistical test for measuring volatility. (Future applications will be to the value shortfall and the apparent premium on equities.)

5.1 Description of the Computational Model

In Section 3, the characteristics of a firm were $f = (i, \ell, k)$. In the present section, all firms will have $\ell = \infty$ and $k = 1$. The firm's opportunity to make an infinitely long-lived commitment is limited to a single date at which it first employs resources. Thereafter the firm simply observes the consequences of its decision.

The set of firms is

$$F = \{(i,t); i = I, II \quad t = \ldots, -2, -1, 0, 1, 2, \ldots\} \cup \{f_o\},$$

where $f = (i,t)$ identifies the sector and date at which the firm makes its commitment.

The unlimited length of the commitment will imply that once a firm is activated, it never becomes inactive. To lessen the weight of the past, we shall assume (contrary to Section 3) that the resource good depreciates in use at a constant exponential rate δ. To prevent the resource good from depreciating to zero, the economy is endowed with a constant amount of *new resources* $\eta > 1$ in each period. Thus, the total amount of resources is

$$w_t = \eta \sum_{\tau=0}^{\infty} \delta^{t-\tau} = \frac{\eta}{(1-\delta)}.$$

In addition to the η new units of resource good available in each period, one unit of *new capacity for adding firms* is available that can be used in either sector. If not fully utilized, it suffers 100% depreciation. This last aspect greatly simplifies the decision problem, since it removes the option of waiting to invest the capacity.

Let $x(t,t) \in [-1, 1]$ denote the choice of sectors in which new firm capacity and resources are added: $x(t,t) = 1$ if capacity and resources are invested in Sector I and $x(t,t) = -1$ if in Sector II at time t. (Note: $x(t,t) = 0$ means that one-half unit of resources is added to each sector.) Once committed to a particular sector, both the capacity and resource become specific and depreciate at the common rate δ. Depreciation implies that

$$|x(t-\tau, t)| = \delta^{t-\tau} |x(t-\tau, t-\tau)|,$$

where $x(t-\tau, t)$ represents the allocation of resources committed to firms at $t - \tau$ remaining at t.

An allocation at t may therefore be described by

$$a_t = \{x(t-\tau, t) : \tau = 0, 1, 2, \ldots\}.$$

For the results below, a_t will be a sequence of 1's and -1's multiplied by their appropriate depreciation factors $\delta^{t-\tau}$.

The total capacity of new firms to employ resources in any period is 1. Hence, at a minimum, the remaining $(\eta - 1)$ of resources available in any period is allocated to f_o. Assume that the riskless payoff d from f_o is positive. Further, assume the representative consumer is risk-neutral. Recalling the results of Section 4.2.2, setting the constant marginal utility of consumption equal to 1, the valuation of the resource good is therefore:

$$R = d\frac{\eta}{1-\delta}\frac{\beta}{1-\beta}.$$

(We switch to ex-dividend values to simplify the description of the econometric tests.)

It remains to describe the productivity shocks. Let

$$s_i(t, t+\tau) = \begin{cases} s + \sum_{n=1}^{\tau} Z_{t+n} & \text{if } i = I \\ s - \sum_{n=1}^{\tau} Z_{t+n} & \text{if } i = II \end{cases}$$

be the shock per unit of resource good allocated. The shock to firms in either sector includes a constant s. The stochastic component says that once resources are committed, their per unit productivity depends on the sum of shocks thereafter. The opposing symmetry follows the previous convention of letting shocks from $Z(1)$ ($\mathbf{E}Z(1) > 0$) favor commitments in sector I and shocks from $Z(2)$ favor sector II. The partial sums of productivity shocks produce heterogeneity within and across firms in sectors I and II in a manner which attempts to mimic the effect of varying commitment length on aggregate quantities and values in the theoretical model.

The goal of the following section is to describe the separate components of wealth,

$$W = V_A + V_{\sim A} + R.$$

5.2 Analysis of the Computational Model

The assumption of risk neutrality of the representative consumer further simplifies the allocation problem since it will allow us to ignore the past in making current decisions.

Assume $s > d$. Since $\mathbf{E}Z = 0$, resources allocated to either sector have a greater expected payoff than f_o. Hence, exactly $\eta - 1$ units of the resource good — the amount left over after allocating resources up to capacity in sectors I and II — will be devoted to f_o. The complexity of the reallocation process has been reduced by fixing the amount of resources available for reallocation each period.

Let

$$\tilde{b}_t = \mathbf{P}[\lambda_t = 1|\omega_t]$$

be the posterior on λ_t given the observed information up to time t and let

$$b_t = \alpha\tilde{b}_{t-1} + (1-\alpha)(1-\tilde{b}_{t-1})$$

which is equal to $\mathbf{P}[\lambda_t = 1|\omega_{t-1}]$, the posterior on λ_t given the observed information up to time $t-1$.

Let $\mathbf{E}Z(1) = \mu$ ($\mathbf{E}Z(2) = -\mu$). Consider, the optimal forecast of $Z_{t+n}, n > 0$ based on information up to time t:

$$\mathbf{E}[Z_{t+n}|\omega_t] \quad = \quad \mu\mathbf{P}[\lambda_{t+n} = 1|\omega_t] - \mu\mathbf{P}[\lambda_{t+n} = 0|\omega_t]$$

$$= \mu\mathbf{P}[\lambda_{t+n} = 1|\omega_t] - \mu(1 - \mathbf{P}[\lambda_{t+n} = 1|\omega_t])$$
$$= 2\mu\Big(\mathbf{P}[\lambda_{t+n} = 1|\omega_t] - 1/2\Big)$$

Assume $\alpha > 1/2$; then if $b_{t+1} > 1/2$, $\mathbf{P}[\lambda_{t+n} = 1|\omega_t]$ converges monotonically to $1/2$ from above and if $b_{t+1} < 1/2$ it converges monotonically to $1/2$ from below, i.e., the optimal forecast of future values of Z converges to the unconditional expection as the future date increases.

THE OPTIMAL POLICY: The risk neutral consumer will prefer to invest all new resources and capacity in a new firm in sector I if $b_{t+1} > 1/2$ and in sector II if $b_{t+1} < 1/2$. This follows directly from the monotonicity property of the expectation of future Z_{t+N} given b_{t+1}.

Let $v_i(t-\tau, t)$ be the value of firm $f = (i, t-\tau)$ at date t, $i = I, II$. It is made up of the following three components:

$$\begin{aligned} v_i(t-\tau,t) &= x(t-\tau,t)\Bigg(\Big[(s-d)\frac{\beta}{1-\beta}\Big] + \Big[\frac{\beta}{(1-\delta)(1-\beta)}\sum_{n=0}^{\tau} z_{t-n}\Big] \\ &\qquad + \Big[\sum_{n=1}^{\infty}\beta^n\Big(\delta^n\mathbf{E}[\sum_{h=1}^{n} Z_{t+h}\,|\,\omega_t]\Big)\Big]\Bigg) \end{aligned}$$

The first component represents the constant productivity advantage of sector I (or sector II) firms over the (marginal) firms of type f_o. The second component represents the value of previous shocks. If the firm is relatively young this component should be positive on average, since b_{t+1} must have been greater than or equal to $1/2$ if $i = I$ and less than or equal to $1/2$ if $i = II$. As the firm ages it is possible that λ will have switched at least once, hence the second component might be small or even negative. The third component represents the value of future productivity shocks to the firm. As the firm ages the sign of this term is indeterminate; however, its value is goes to zero as $(t-\tau)$ increases.

The total value of all *active firms* at t is

$$V_A(t) = \sum_{\tau=0}^{\infty}\Big[v_I(t-\tau,t) + v_{II}(t-\tau,t)\Big]$$

It can be similarly be split into three components:

Value of Constant Productivity Advantage over $\{f_o\}$

$$(V_{A1}) \qquad (s-d)\frac{\beta}{1-\beta}$$

Value of Previous Shocks to Active Firms

$$(V_{A2}) \qquad \frac{\beta}{(1-\delta)(1-\beta)}\sum_{\tau=1}^{\infty}\Big(x(t-\tau,t)\cdot\sum_{n=1}^{\tau} z_{t-n}\Big)$$

Value of Future Shocks to Active Firms

$$(V_{A3}) \qquad \bar{a}_t \sum_{n=1}^{\infty} \beta^n \left(\delta^n \mathbf{E}[\sum_{\tau=1}^{n} Z_{t+\tau} \,|\, \omega_t] \right).$$

where

$$\bar{a}_t = \sum_{\tau=1}^{\infty} x(t-\tau, t),$$

is the net difference between the allocation of resources at t previously allocated to sector I over sector II.

Items V_{A1} and V_{A2} comprise the aggregate dividends produced by active firms, therefore aggregate dividends. The sum of these three components does not equal the total value of *all* firms in the economy since new firms will appear in the future. These new firms are the equivalent of the currently inactive firms of the theoretical model that are expected to reappear. The value of new firms in the computational model is

$$V_{\sim A}(t) = \sum_{\tau=t+1}^{\infty} \beta^{\tau-t} \left(\sum_{n=1}^{\infty} \beta^n \mathbf{E}[x(\tau, \tau+n) \sum_{i=\tau}^{n} Z_i \,|\, \omega_t] \right)$$

The valuation depends on the future (optimal) decision of where to allocate capacity.

Aggregate dividends represent the flow from currently active firms but the characteristics of these firms are constantly changing through the addition of new firms and the depreciation of old ones. For there to be a value shortfall, i.e., $V_{\sim A} > 0$, future allocation decisions must be based on more than the unconditional values of future Zs. For example, if $\alpha = 1/2$, previous values of Z are useless in predicting future values. Hence, after subtracting out the constant rent, the marginal expected value of firms activated in the future would be zero. When $\alpha > 1/2$ the situation is different: as α gets closer to one, or μ becomes larger, or σ becomes smaller, the value of $V_{\sim A}/V_A$ will increase. Each of these conditions implies that previous values of Z are useful for future decisions and therefore contributes to increasing the value shortfall.

5.3 The West Volatility Test

We simulate a calibrated version of the computational model to assess the potential bias in a canonical test for excess volatility due to West (1988). His test is designed to deal with the 'stationary issue' which had cast some doubt on the results from the Shiller-LeRoy-Porter tests. West's test is based on an inequality produced by linear predictions from two sets of information, one a subset of the other. The intuition is that the revision in the forecast of the stock price will be smaller, the larger the information set used to produce it.

Recall that the value of existing firms on the stock market depends on three components, two of which are know at the start of time t, V_{A1} and V_{A2}. Thus, for existing firms the only unanticipated movement in stock prices must come from the realization z_t and the manner in which this affects future forecasts of the partial sums of Z_{t+n}.

Consider, the update in the optimal forecast of $Z_{t+n}, n > 0$ produced by the realization of Z_t:

$$\mathbf{E}[Z_{t+n} \mid \omega_t] - \mathbf{E}[Z_{t+n} \mid \omega_{t-1}] = 2\mu(\tilde{b}_t - b_t)(\mathbf{P}[\lambda_{t+n} = 1 \mid \lambda_t = 1] - \mathbf{P}[\lambda_{t+n} = 1 \mid \lambda_t = 0]).$$

The only source of time variation in this expression is the update in the posterior $\tilde{b}_t - b_t$.

Define

$$J_t(\omega_t) = \sum_{n=1}^{\infty} (\delta\beta)^n \mathbf{E}[\sum_{\tau=0}^{n} Z_{t+\tau} \mid \omega_t]$$

as the present discounted value of the partial sums of predicted Z_{t+n}. Using the change in the optimal forecast of Z_{t+n}, the unanticipated movement in J_t is

$$U_t(\omega_t) = J_t(\omega_t) - \mathbf{E}[J_t \mid \omega_{t-1}]$$

where

$$J_t - \mathbf{E}[J_t \mid \omega_{t-1}] = \sum_{n=1}^{\infty} (\delta\beta)^n \Big((Z_t - \mathbf{E}[Z_t | \omega_{t-1}]) + \sum_{\tau=1}^{n} \mathbf{E}[Z_{t+\tau} \mid \omega_t] - \mathbf{E}[Z_{t+n} \mid \omega_{t-1}] \Big).$$

The unanticipated changes in future values of Z is a *non-linear* function of Z_t. Information about ω_{t-1} can be summarized by b_t, so by abuse of notation we can also write $U(Z_t, b_t)$.

Before we consider the effect of reallocation, we need to address two issues concerning the testing methodology applied to U. The caveats we note in this case are also important for the application of excess volatility tests to individual equities where the effect of reallocation can be ignored.

To produce a forecast of the fundamental value of J from a smaller information set, the econometrician must first construct a forecasting model for the future values of the sums of Z_t and then estimate it using the observed values of the partial sums.

and then estimate it for partial sums of Z_t based on their previous history. Predictions from the forecasting model are used to form a 'fundamentals' estimate of J_t and the unanticipated change U_t. In practice the forecasting model is always assumed to be linear. Further, it is generally assumed (with exceptions) that agents in the economy are not learning and therefore the forecast errors to be used in generating the unanticipated movement are the in-sample residuals from the linear forecasting model. *Neither of these assumptions are satisfied the computational model.*

5.3.1 Nonlinearity

West's (1988) proof of the consistency of his test is only valid under the assumption that the variable being forecasted is a linear time series. One might think that if the forecasted variable were non-linear that this would only affect the power of the test rather than its size, i.e., if the best forecasting model is really nonlinear, restricting attention to linear models is similar to using a smaller information set. Thus, the bound is still working in the correct direction under the null hypothesis. As discussed in Potter (1994) this intuition is correct when applied to the one-step-ahead forecast error $(Z_t - \mathbf{E}[Z_t \,|\, \omega_{t-1}])$, but not when applied to $(\mathbf{E}[Z_{t+n} \,|\, \omega_t] - \mathbf{E}[Z_{t+n} \,|\, \omega_{t-1}])$. In fact there appears to be no general relationship between sequences of differences in nonlinear predictors based on different information sets and sequences of differences in linear predictors based on the same information sets as the nonlinear predictor.

5.3.2 Use of full sample information

In the computational model, the agents know the parameters of the forecasting model, (α, μ, σ), but they do not know the current value of λ. The agents generate recursive forecasts that do not react to improved knowledge of the historic values of λ: 'byegones are byegones' as far as agents are concerned. Econometric forecasting models that are estimated on the whole sample without constructing a recursive estimate of beliefs will react differently to this information about past values of λ. The reaction will be to produce an ex-post smooth estimate of the new information about dividends in the past. In particular, linear models estimated over the whole sample that ignore the non-observability of λ will frequently lead to better predictions about dividends over the sample period than could be obtained by agents based on their optimal (non-linear) estimates recursively applied. In other words, the joint misspecification of linearity and observability will lead to the conclusion that agents had better information at the time than they did. Hence, the hypothesis upon which West relies— that the econometrician's information set is smaller than the agents'—can lead to the conclusion that the econometrician knows more than the agents when applied under the joint misspecification.

5.3.3 Application to the Model

The unanticipated movement in the value of active firms is the product of U and the summary statistic $\bar{a}$ of the current allocation,

$$(V_A)_t - \mathbf{E}_{t-1}(V_A)_t = U_t \bar{a}_t.$$

When $\bar{a}_t = 0$, the distribution of a_t is such that revisions concerning future Z_{t+n} will cancel; firms experiencing positive revaluations will be exactly offset by firms experiencing negative revaluations. However, as the absolute value of $\bar{a}_t$ gets closer to its maximum of $1/(1-\delta)$, the affect of unanticipated changes will be greatly amplified.

Suppose the econometrician is told the sequence $\{U_t \bar{a}_t\}$ (West suggests an instrumental variable estimator in practice). The objective is then to compare the variance of $U\bar{a}$ with a value derived from a forecasting model for aggregate dividends, call this $FU\bar{a}$. Since the information set produced by aggregate dividends is presumed smaller than that used in the construction of $U\bar{a}$ in large samples or repeated sampling, the variance of $FU\bar{a}$ will be larger than $U\bar{a}$.

In the computational model, aggregate dividends are produced by a changing sample of firms. By construction this will tend to be smoother than the dividend stream produced by any particular set of firms. We assume that the econometrician ignores this feature. In addition the econometrician

- ignores the fact that U is not linear
- uses the whole sample for estimation of the forecasting model.

We call the econometric model of unanticipated movements in firm valuations produced in this way the linear decomposition, LD_t.

West (1988) reports a Monte Carlo study in the case that dividends (D) follow an arithmetic random walk

$$D_t = D_{t-1} + \varepsilon_t,$$

where ε_t is iid. Therefore, in our terminology his $U\bar{a} = V - \mathbf{E}V = \varepsilon(\beta/(1-\beta))$. He examines the following ratio:

$$\text{WEST} = -100\frac{\sigma^2(LD) - \sigma^2(U\bar{a})}{\sigma^2(U\bar{a})},$$

where $\sigma^2(U\bar{a})$ is the variance of the true unanticipated movement in the value of active firms and $\sigma^2(LD)$ is the variance of the unanticipated movement obtained by applying a present value calculation to the forecast error from the linear model for aggregate dividends.

By construction the present value relationship holds and one would expect the median value of the statistic to be 0 if the information set used was the same as market participants. West finds some mild small sample bias in his Monte Carlo study with the median estimate being 1.7. In the data he finds a value of 80.

We use the same procedure except that we replace the arithmetic random walk process for dividends by a calibrated version of the computational model. Dividends, stock price values and the unanticipated movement were simulated with the discount factor set at 0.9413 and $\delta = 0.95$ and the change in dividends calibrated to the coefficient of variation given in West (1988). Notice that one has great freedom in choosing s, d, η since they do not enter into the definition of $U\bar{a}$. For example, one could introduce growth through deterministic productivity improvements in s and d; and other features such as price dividend ratios could be matched by changing s relative to the properties of Z_t. The ratio of dividends to national income could be matched by choice of η. In the calibration s_t and d_t are assumed to grow at the same rate (with a constant level difference set) to match the coefficient of variation of changes in dividends.

For each sample of aggregate dividends a linear autoregressive model with 40 lags was estimated on the whole sample and used to construct a hypothetical upper bound on the volatility of the true unanticipated movement in stock market value.

The first column of Table 1 replicates West's Monte Carlo study for a sample size of 140, which gives an effective sample of 100 (1,000 replications were used). The median value of WEST is close to that found in the data even though by construction the present value model is true (33% of the replications had a greater value than West found in the data). There is substantial variation in the values as can be seen from the maximum and minimum of WEST.

The other two columns of the table show the effect of increasing the effective sample size to 1,000 (100 replications) and 10,000 (10 replications) respectively. Note that in these two columns the minimum is always above zero and the median appears to be settling to a value close to 50. In West's terminology an econometrician in the economy with 10,000 years of stock market data would think that 50% of the volatility in stock prices was excessive.

5.3.4 Other Statistical Properties of the Simulation

The correlation between the true unanticipated price movement and that implied by the *linear* model for dividends was also calculated. Its median value across simulations was close to zero. Along with the results of the volatility test, this might be offered as supporting evidence that the stock market is not being driven by 'fundamentals.' In fact, the absense of correlation is based on the application of a linear model in a non-linear environment.

The properties of dividends in this economy are also of interest. Although by construction dividends are integrated of order zero (on removal of the deterministic trend), the sum of the moving average coefficients from the linear autoregressive model estimated on dividend growth had an average value close to 1. This is reflective of the 'nearly integrated' nature of the productivity shocks.

Finally, one might wonder whether the nonlinearity is so obvious that an econometrician would immediately reject the linear model. One measure of this propensity is the standard deviation of the one-step ahead forecast error for dividends from the linear model estimated over the whole sample compared to the optimal forecast error using the nonlinear predictor. In sample sizes of 100, the linear forecast always had a smaller standard deviation than the true error. The explanation is that the linear predictor smooths based on the whole sample (both revised views of λ and the allocation) so it can avoid 'mistakes' made by the optimal predictor. In sample sizes of 10,000 this effect disappears since the number of autoregressive coefficients is kept constant at 40. However, the average improvment from the optimal predictor is only 1.5% in the sample of 10,000; so there appears to be little reason for rejecting linearity.

Figures 6.a to 8.b give information from three samples on (1) the true unanticipated price movement, (2) the price movement implied by the linear decomposition and (3) the pattern of beliefs compared to the true value of λ_t. Volatility in the true unanticipated component of prices appears to be tightly related to sudden movements in the prior, reflecting the re-valuation of existing fixed commitments. In contrast the unanticipated movement using the linear model for dividends is much smoother and more connected to the true movement in λ_t, reflecting the full sample smoothing implicit in the linear model. The unanticipated movements are somewhat asymmetric with large negative movements ocurring frequently.

Notice that large movements in beliefs do not always produce big movements if in the previous periods beliefs had not been consistently favoring one sector over the other imply $a_t \approx 0$ (see Figure 7 around observation 25).

References

[1] W.A. Brock. *Asset Pricing in a production Economy.* University of Chicago Press, Chicago, 1982.

[2] A. David. Fluctuating confidence and stock-market returns. Technical report, Board of Governors of the Federal Reserve, 1994.

[3] A. David, S. Oh, J. Ostroy, and K. Shin. Cyclical fluctuations in uncertainty: An application to the premium on equities. Technical report, Department of Economics, UCLA, 1992.

[4] L. Epstein and S. Zin. First order risk-aversion and the equity premium puzzle. *Journal of Monetary Economics*, 26:387–407, 1990.

[5] S.J. Grossman and R.J. Shiller. The determinants of the variability of stock prices. *American Economic Review*, 71:222–27, 1981.

[6] B. H. Hall and R. E. Hall. The value and performance of u s corporations. *Brookings Papers on Economic Activity*, (1):1–49, 1993.

[7] J.D. Hamilton. A new approach to the economic analysis of economic time series and the business cycle. *Econometrica*, 57:357–384, 1989.

[8] L.P. Hansen and R. Jagannathan. Implications of security market data for models of dynamic economies. *Journal of Political Economy*, 99:225–262, 1991.

[9] H. Hopenhayn. Exit, selection and the value of firms. *Journal of Economic Dynamics and Control*, 16:621–653, 1992.

[10] R.A. Jones and J.M. Ostroy. Flexibility and uncertainty. *Review of Economic Studies*, LI:13–32, 1991.

[11] S.F. LeRoy and R. Porter. The present value relation: Tests based on implied variance bounds. *Econometrica*, 49:555–574, 1981.

[12] R.E. Lucas. Asset prices in an exchange economy. *Econometrica*, 46:1429–1445, 1978.

[13] R. Mehra and E.C. Prescott. The equity premium: A puzzle. *Journal of Monetary Economics*, 15:145–61, 1985.

[14] J. Olley and A. Pakes. The dynamics of productivity in the telecommunications equipment industry. Technical report, Department of Economics, Yale University, 1991.

[15] S.M. Potter. Nonlinear impulse response functions. Mimeo, Department of Economics, UCLA, 1994.

[16] R. Shiller. Do stock prices move too much to be justified by subsequent changes in dividends. *American Economic Review*, 71:421–36, 1981.

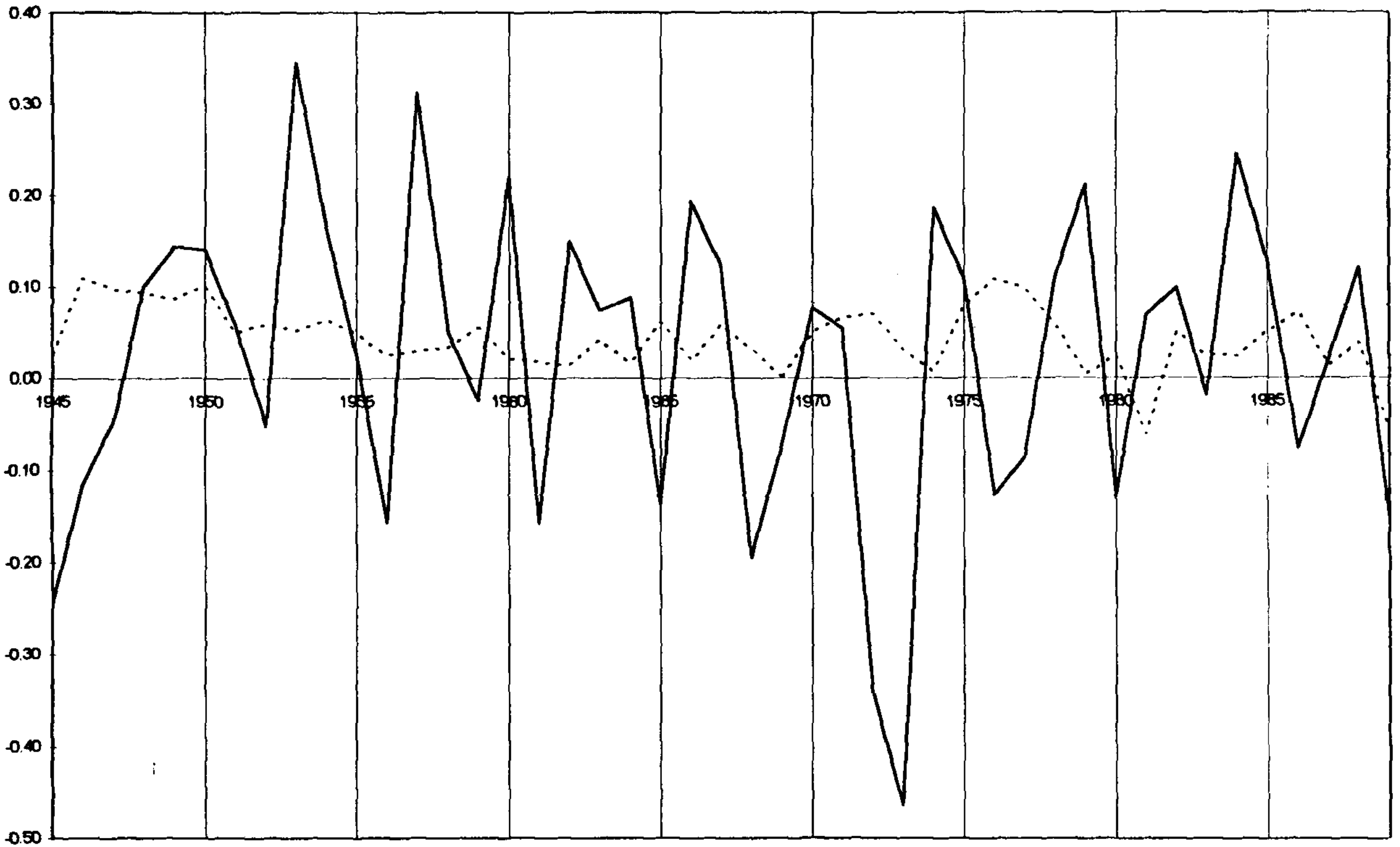
[17] K. D. West. Dividend innovations and stock price volatility. *Econometrica*, 56:37–61, 1988.

Table 1:

Monte Carlo Study of West Volatility Statistic using the Computational Model

Sample Size	100	1,000	10,000
MEDIAN	72	50	48
MAXIMUM	99	65	52
MINIMUM	-1563	24	43

National Wealth: Owner-Occupied Real Estate and Corporate Equity*

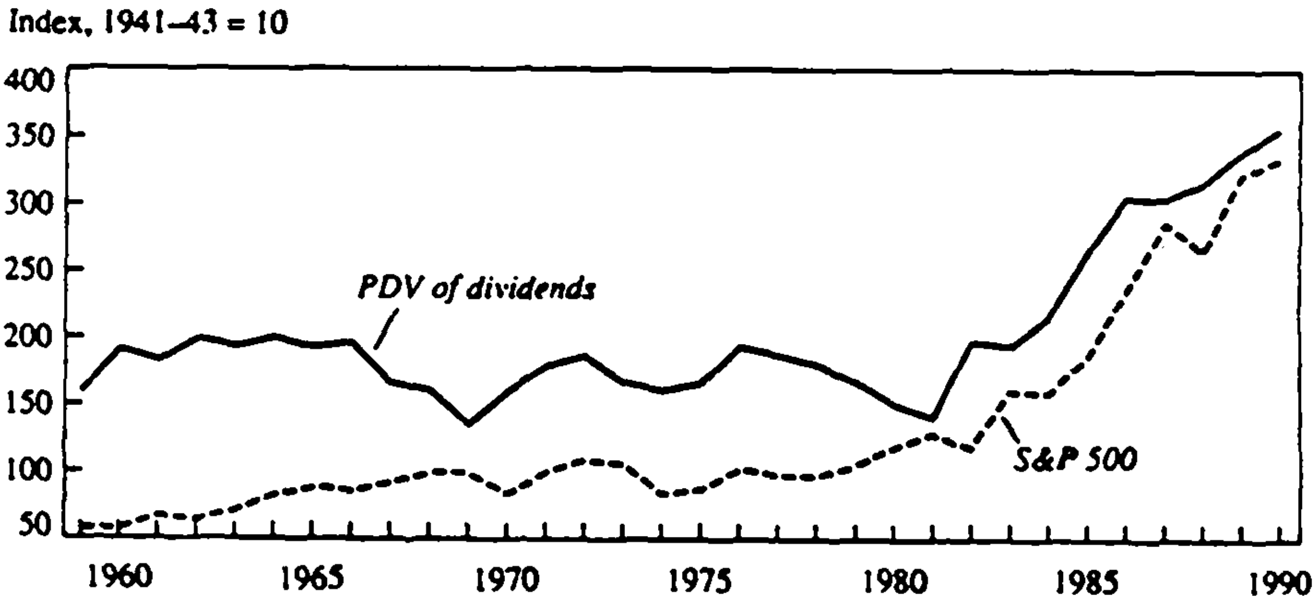


*SOURCE: 1992 Economic Report of the President, Table B-110.

Figure 1

FIGURE 2

Figure 1. The S&P 500 Stock Price Index and the Risk-free Present Discounted Value of Its Future Dividends



Source: *Economic Report of the President*, 1992, table B-91.

Figure 3:
Value Shortfall in CRSP

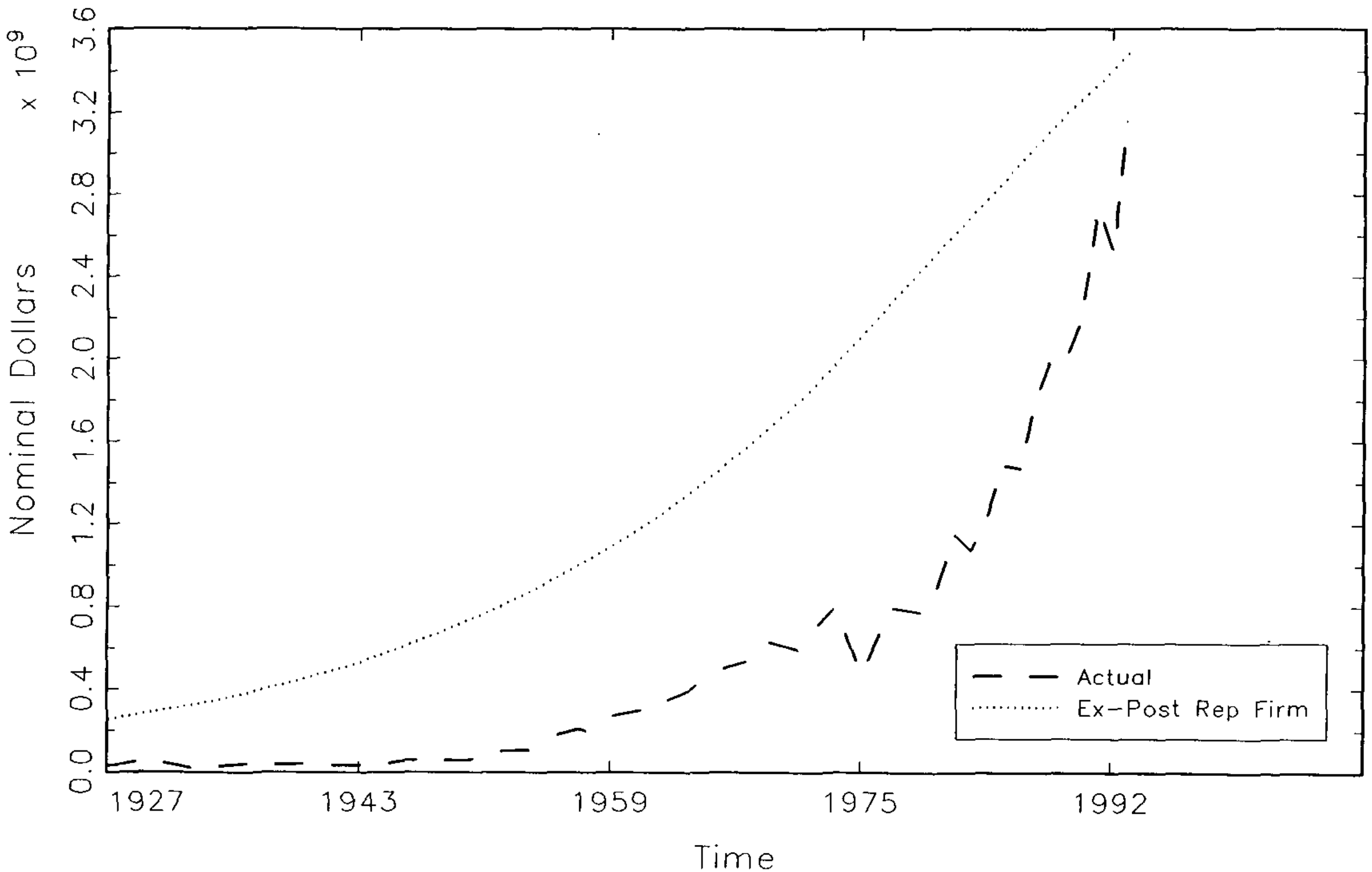


Figure 4:
Stochastic Discount Factor for
Constant Relative Risk Aversion Utility

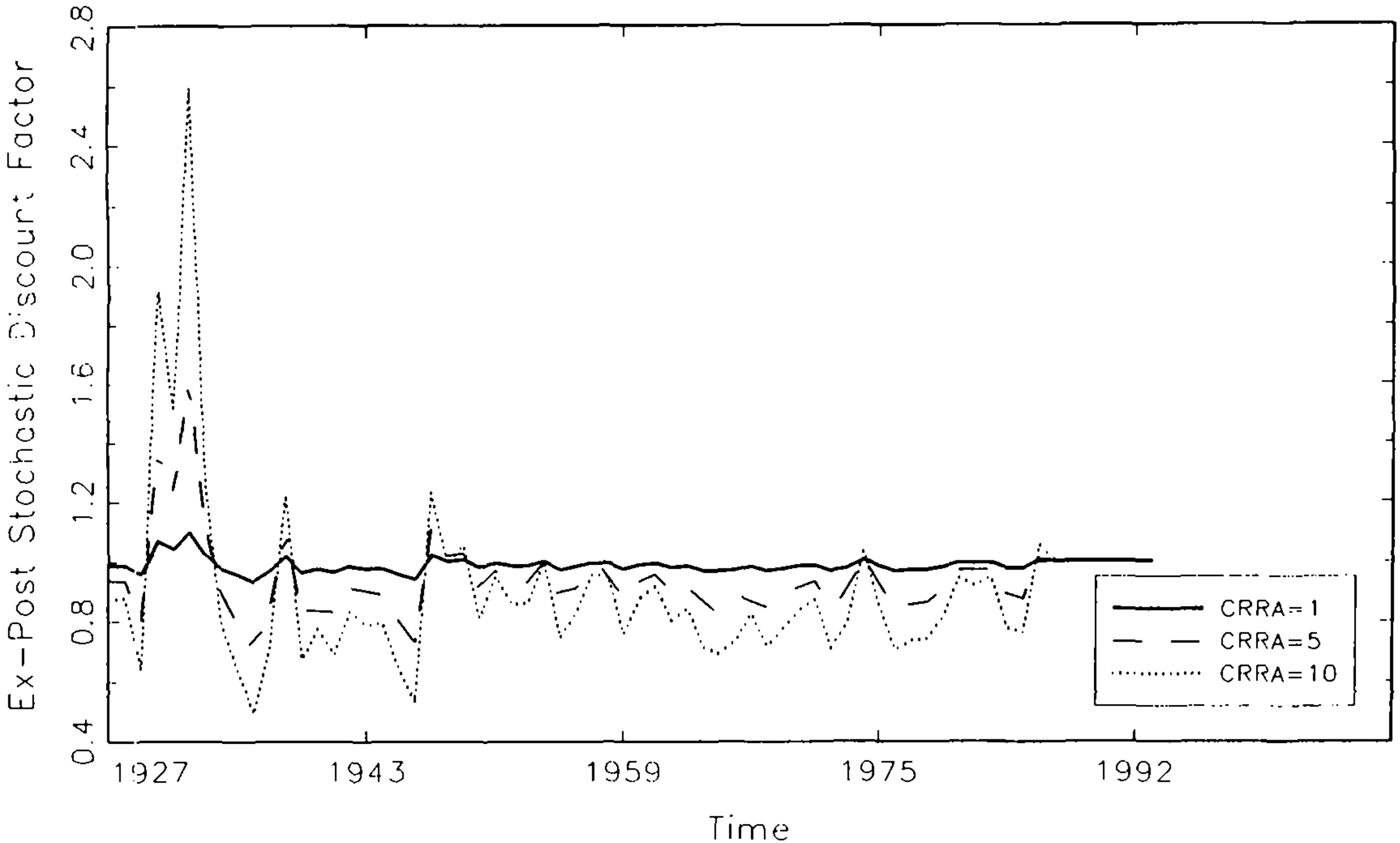
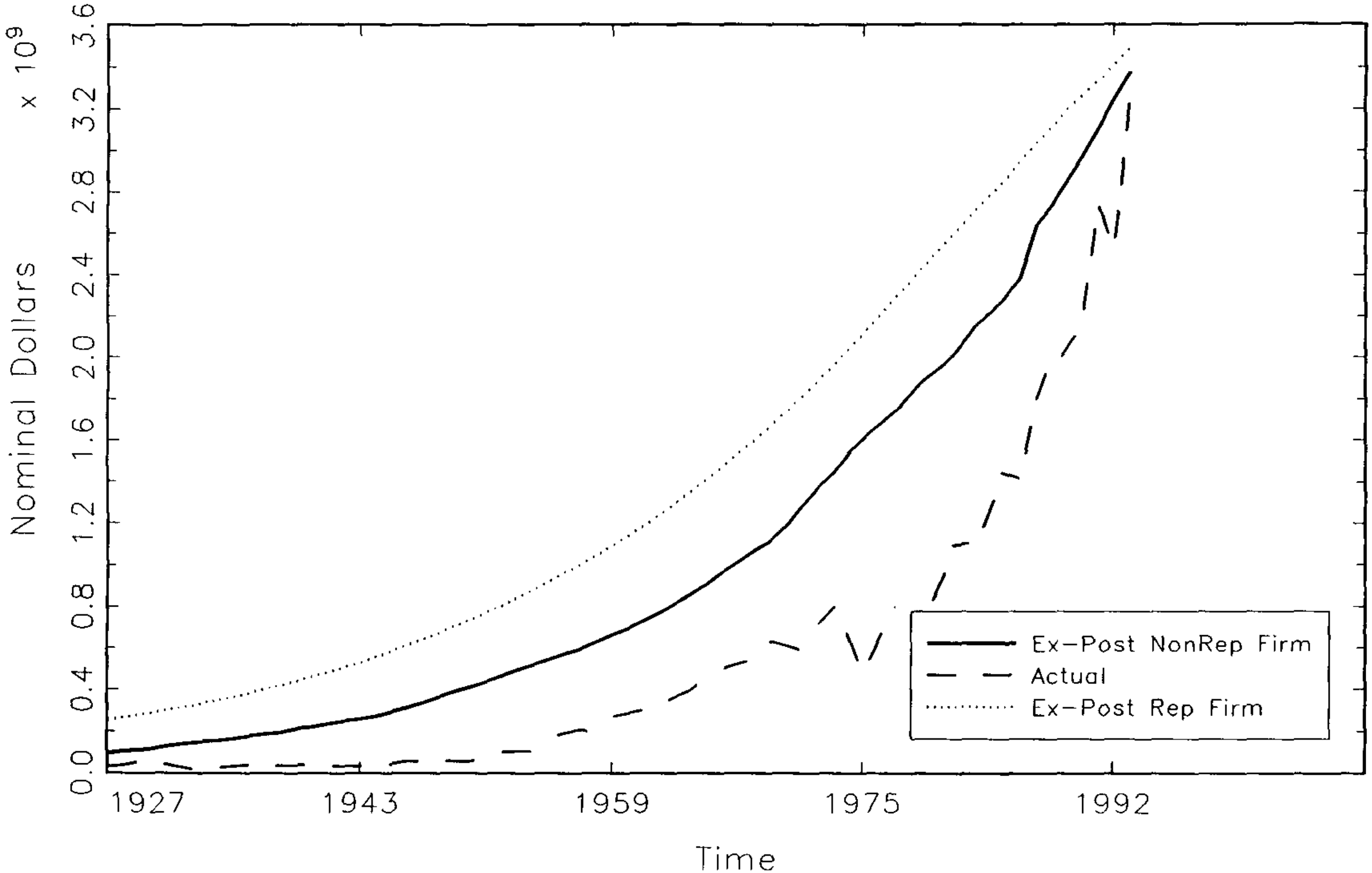
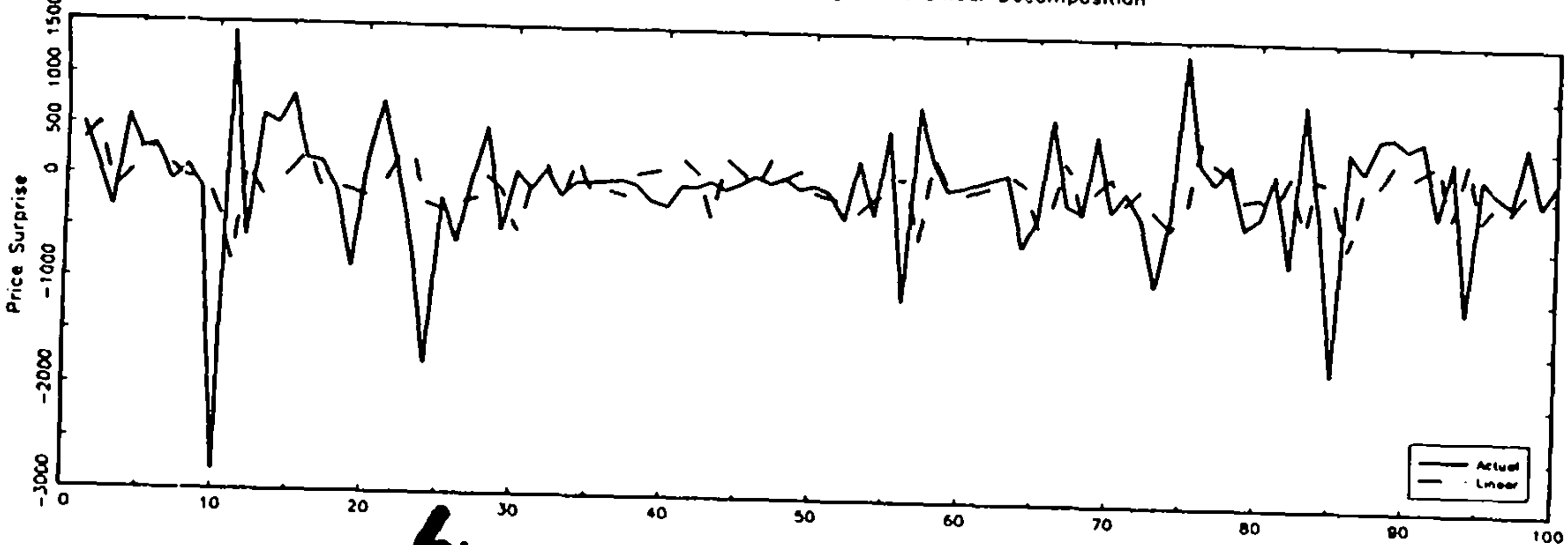


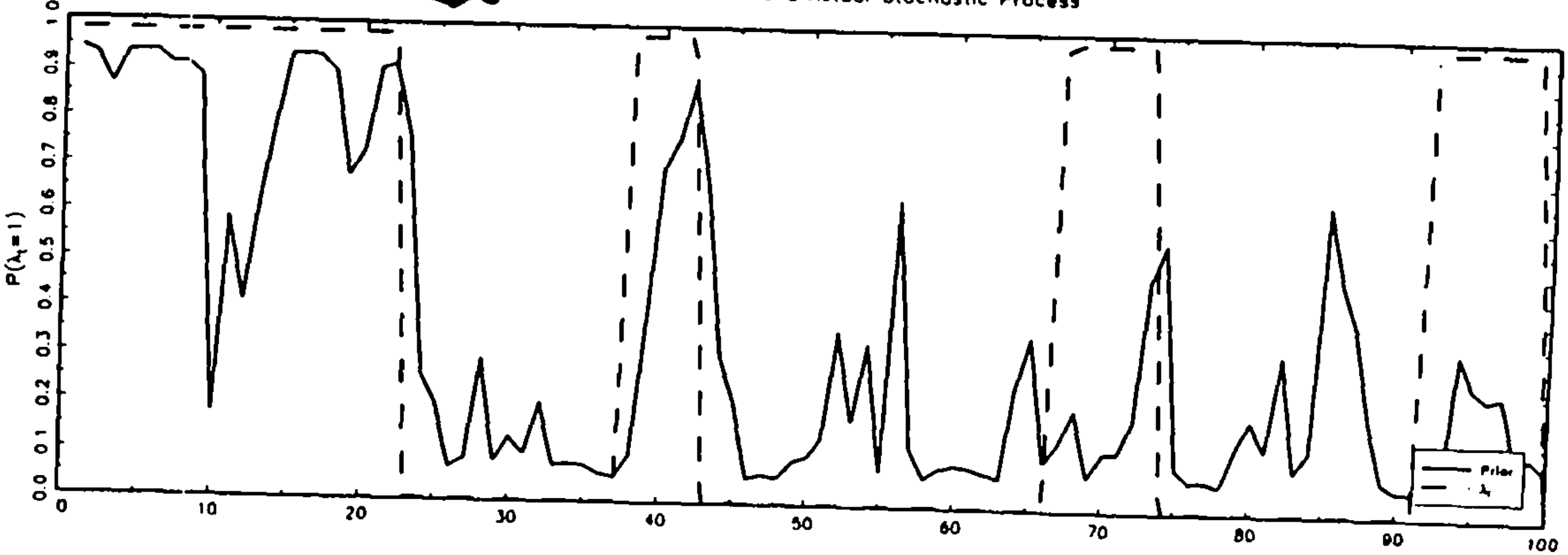
Figure 5:
Value Shortfall: Year by Year CRSP



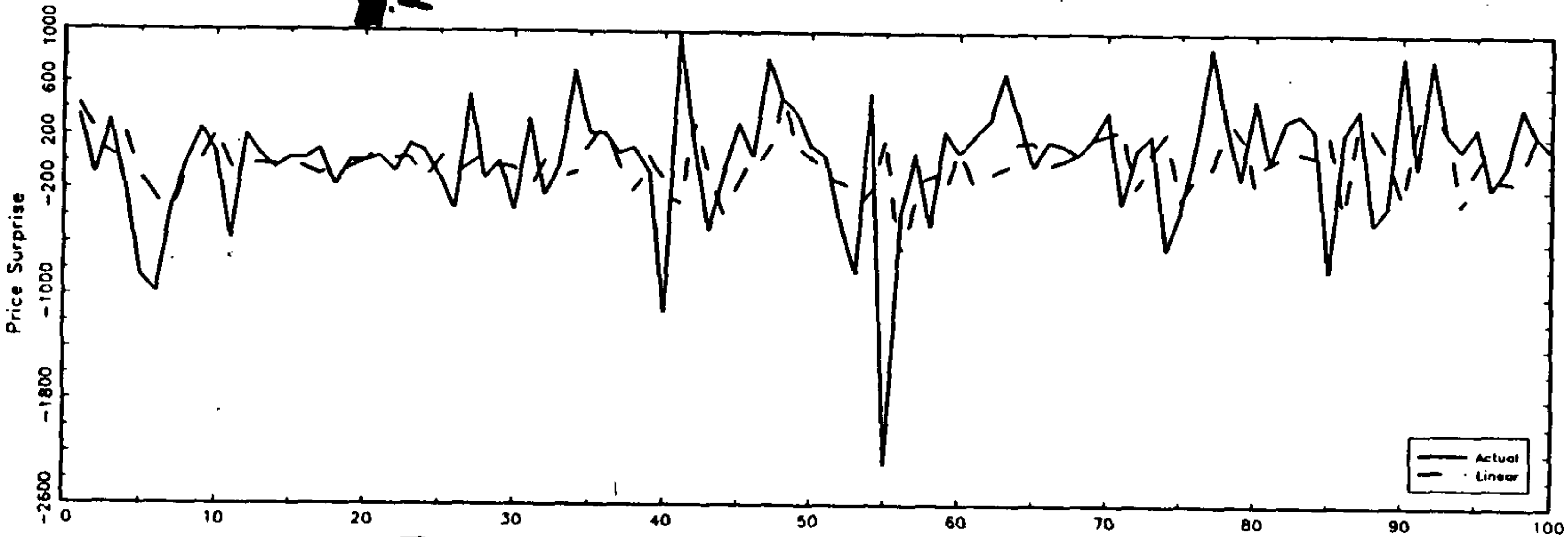
6a Actual Unanticipated Price Changes and Linear Decomposition



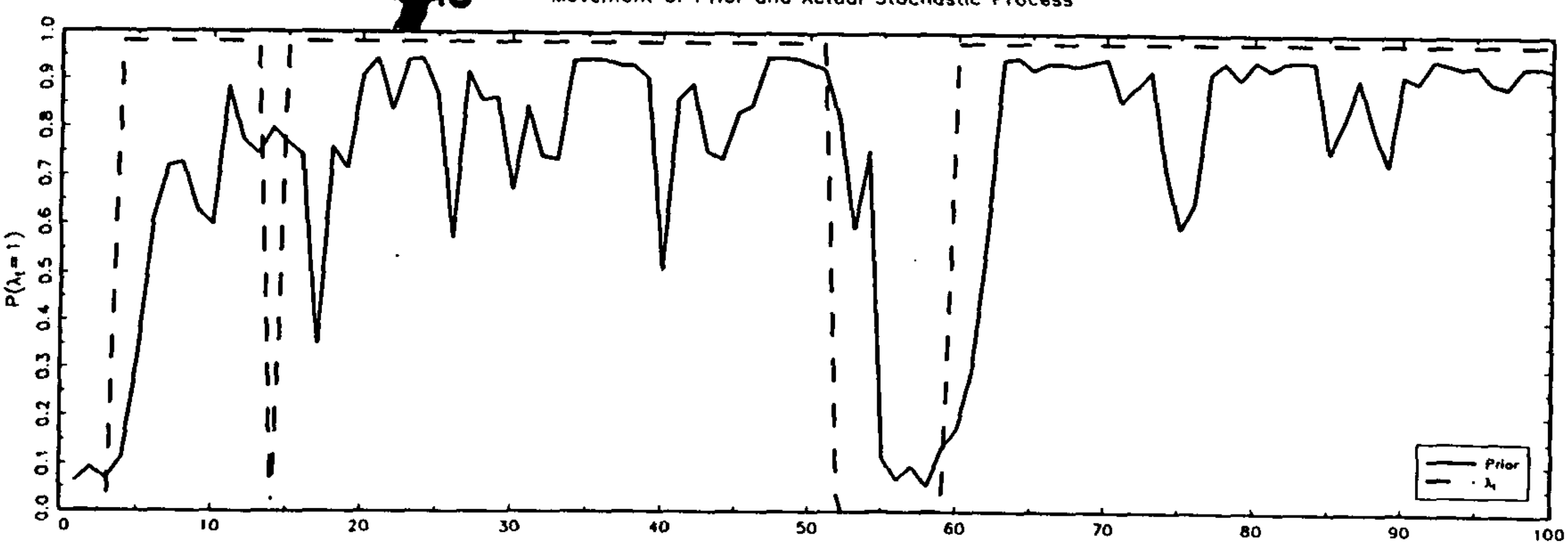
6b Movement of Prior and Actual Stochastic Process



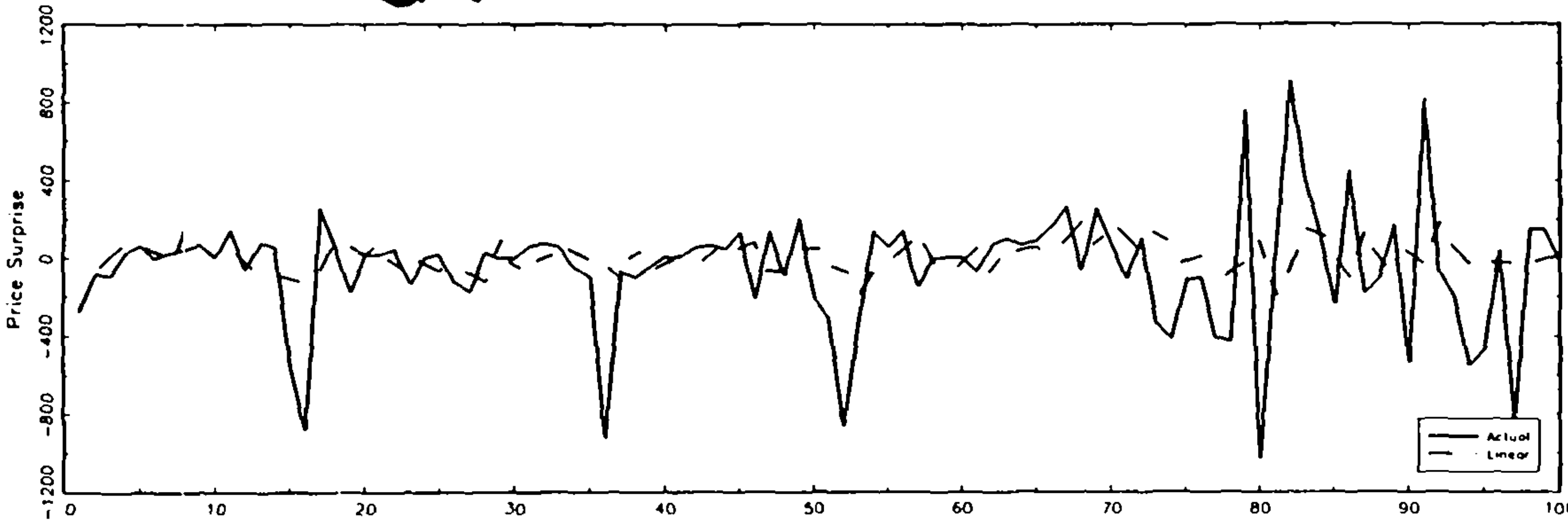
7a Actual Unanticipated Price Changes and Linear Decomposition



7b Movement of Prior and Actual Stochastic Process



8.a Actual Unanticipated Price Changes and Linear Decomposition



8.b Movement of Prior and Actual Stochastic Process